

## DOCUMENT RESUME

ED 132 046

SE 021 723

AUTHOR Karplus, Robert; And Others  
TITLE Proportional Reasoning and Control of Variables in Seven Countries. Advancing Education Through Science-Oriented Programs, Report ID-25.  
INSTITUTION California Univ., Berkeley. Lawrence Hall of Science.  
SPONS AGENCY National Science Foundation, Washington, D.C.  
PUB DATE Jun 75  
NOTE 62p.; Contains occasional light type  
EDRS PRICE MF-\$0.33 HC-\$3.50 Plus Postage.  
DESCRIPTORS \*Cognitive Development; Cognitive Measurement; Cognitive Tests; \*Educational Research; \*Evaluation; \*Science Education; Secondary Education; Secondary School Science; \*Student Evaluation  
IDENTIFIERS Europe; \*Piaget (Jean); Research Reports; United States

## ABSTRACT

Reported are the results of a study of the Piagetian formal thought range of 3600 students, 13 to 15 years old, in seven countries: Denmark, Sweden, Italy, United States, Austria, Germany, and Great Britain. Cognitive measurement was obtained by group administration of two tasks to assess proportional reasoning and control of variables. Overall results found 25% of the students at the formal level, 32% of the students in a transition level, 15% of the students at an additive level, and 28% of the students at an intuitive level. Scoring procedures and summary results for each country are also included. (SL)

\*\*\*\*\*  
\* Documents acquired by ERIC include many informal unpublished \*  
\* materials not available from other sources. ERIC makes every effort \*  
\* to obtain the best copy available. Nevertheless, items of marginal \*  
\* reproducibility are often encountered and this affects the quality \*  
\* of the microfiche and hardcopy reproductions ERIC makes available \*  
\* via the ERIC Document Reproduction Service (EDRS). EDRS is not \*  
\* responsible for the quality of the original document. Reproductions \*  
\* supplied by EDRS are the best that can be made from the original. \*  
\*\*\*\*\*

PROPORTIONAL REASONING AND CONTROL OF VARIABLES

IN SEVEN COUNTRIES

by

Robert Karplus\*  
AESOP<sup>+</sup>, Lawrence Hall of Science  
University of California, Berkeley

Elizabeth Karplus  
Campolindo High School  
Moraga, California

Marina Formisano  
University of Rome

Albert-Christian Paulsen  
Royal Danish College  
of Educational Studies  
Copenhagen

(June, 1975)

Lawrence Hall of Science  
University of California  
Berkeley, California 94720

\*John Simon Guggenheim Fellow, 1973-74, and Visiting Professor,  
Division for Study and Research in Education, M.I.T.

<sup>+</sup>This study was supported in part by a grant from the National  
Science Foundation.

U.S. DEPARTMENT OF HEALTH,  
EDUCATION & WELFARE  
NATIONAL INSTITUTE OF  
EDUCATION

THIS DOCUMENT HAS BEEN REPRO-  
DUCED EXACTLY AS RECEIVED FROM  
THE PERSON OR ORGANIZATION ORIGIN-  
ATING IT. POINTS OF VIEW OR OPINIONS  
STATED DO NOT NECESSARILY REPRESENT  
OFFICIAL NATIONAL INSTITUTE OF  
EDUCATION POSITION OR POLICY

The wide diversity of interests and abilities encountered among secondary school students is well known to teachers. The present study was intended to illuminate this range in regards to logical-mathematical reasoning, or what Piaget has called formal thought.<sup>1</sup> We developed two tasks to assess proportional reasoning and control of variables. On these tasks, which could be administered to classroom groups in several countries, the students were asked to solve certain problems and then explain or justify their answers. The logical-mathematical reasoning was exhibited by the form and content of the explanations.

Both areas of thought we selected for investigation are vitally important in science instruction. Proportions play a key role in quantitative relations in science. The idea of a controlled experiment, "keeping all other things constant" to isolate the effect of one variable, is essential for determining unambiguous cause-and-effect relationships in science or other subjects. In view of the general ferment in science education, with new course materials and new teaching methods being introduced into secondary schools on an increasing scale, it seemed worthwhile to investigate students' logical-mathematical reasoning at the present time.

Traditionally, reasoning has been investigated by means of clinical interviews in which a subject confronts a phenomenon or a problem and is asked to make a prediction, explain what he observes, or find a solution.<sup>1</sup> The subject is then asked to justify his reasoning, and is pressed by the interviewer to provide additional reasons or alternate theories regardless of his first response. The resulting conversation is eventually analyzed for the types of relationships, "convincing" arguments, and resolution of ambiguities employed by the subject.

Written responses to a group test clearly do not provide the same depth of information about a single individual. There are many reasons why an individual's first response--and that is what he is likely to write down--may not tap all his intellectual resources. Copying from other students and lack of interest also influence the results. Nevertheless, the written task does have an advantage also, in that it permits rapid surveys of large numbers of subjects internationally. At the same time, much of a student's school work is more closely similar to the written task situation than to the clinical interview. Hence, student performance on the written task may help a teacher diagnose learning problems or recognize able students that need additional challenges.

Both areas we chose have been investigated through clinical interviews and through group tests in the last ten years,<sup>2-13</sup> though no studies that included several national groups have come to our attention. The results of these investigations showed that few average subjects used proportional reasoning unaccompanied by physical actions before age fifteen. Proportional reasoning has therefore been regarded as one of the later components of formal thought to be acquired by adolescents. Control of variables was found to have roots in children's concept of a "fair" experiment that has been noted during the elementary school years,<sup>9,14</sup> but the careful analysis and design of experiments has been a source of difficulty for many high school students and adults.<sup>12,13</sup>

A large-scale international study of science education sponsored by the IEA was completed recently.<sup>15</sup> The tests that were used focused on traditional

content areas of science courses at various educational levels and used large numbers of multiple-choice items for the assessment. Some of these involve controlling variables or proportional reasoning,<sup>16</sup> and we considered gleaned some data on student reasoning from the results. Unfortunately, individual multiple-choice items without student justification can give conclusions about the respondent's reasoning only very indirectly, through a comparison of incorrect choices. We were not able to learn anything beyond the fact that these items had relatively low success rates. We therefore consider our study to complement the IEA study and not to duplicate it.

In this article, we shall first describe the student subjects and then our general method of presentation of the tasks. There follows a detailed account of each task and the scoring procedure used to evaluate the responses. Next, we present an overview of the results in seven countries and comparisons among the populations. We conclude with a more general discussion of our observations and some implications.

#### THE SUBJECTS

This report presents an overview of the results of our study of 13- to 15-year-old students in seven countries: Denmark (Copenhagen area), Sweden (Gothenburg area), Italy (Rome), United States (northeast and north central states), Austria (Vienna), Germany (Göttingen), and Great Britain (London). Almost 1800 boys and an equal number of girls participated. The testing usually occurred in a science or mathematics class, but sometimes a scheduling problem led to the use of other classes. As much as possible, the groups in each school were chosen so as to reduce the effects of any tracking by ability within the school.

To have subjects of closely similar ages in all countries in spite of the differences in school entrance criteria, we worked with eighth graders in the United States, Italy, Austria, and Great Britain, seventh graders with a few eighth graders in Denmark and Sweden, and ninth graders in Germany.\* In Austria and Italy, many children leave school after the eighth grade, so that this age group constituted the oldest group we could sample reliably through school visits.

School organization in the seven countries varies considerably, and we tried to choose the populations for testing to get a fair survey appropriate to each school organization. We do not claim to have a representative sample for any country or even any town, since the duration of our visits, our access to various schools, and our knowledge of the composition of the entire population were completely inadequate for that. Except for Denmark, where the data were quite uniform from community to community, we report results only for the specific groups we tested and not for the entire country or region. In Denmark, Sweden, Italy, and United States, where the schools are comprehensive at the seventh- and eighth-grade levels, we have classified the population groups according to the socioeconomic level

---

\*A special change in the school calendar in 1966 eliminated entering German first graders in the fall of that year. Children who entered school in spring of 1966 were designated as ninth graders at the time of our visit.

of the neighborhood served by the school (derived roughly from parental occupation, education, and/or quality of housing). In Austria, Germany, and Great Britain the schools are divided into types with differing entrance requirements, curricula, and school-leaving ages, and these types served as basis for our classification. Further details about the populations will be given in the sections reporting the data for the individual countries.

Table 1 presents an enumeration of the student samples. These include the principal research groups of seventh to ninth graders, and additional small pilot groups in other grades that we used to gain comparative information about the reasoning of students of differing ages. Because of their small size, we shall not quote the results for these groups but use them only as they give perspective regarding the principal data.

The American reader may be interested in a few other details of school practice in the European countries visited. First, the children's studies beyond the fifth or sixth grade are departmentalized. Second, the children do not have electives, but participate in all their courses as members of the same group. Third, the groups tend to stay together for several years, and also tend to be taught by the same set of teachers for several years.

Table 1 Student Populations and Mean Ages

| Country       | Population       | Principal Group |      |      |      | Pilot Group |      |      |      |
|---------------|------------------|-----------------|------|------|------|-------------|------|------|------|
|               |                  | Girls           |      | Boys |      | Girls       |      | Boys |      |
|               |                  | N               | Age  | N    | Age  | N           | Age  | N    | Age  |
| Denmark       | Middle Class     | 207             | 14.0 | 192  | 14.1 |             |      |      |      |
| Sweden        | Middle Class     | 47              | 14.6 | 51   | 14.5 |             |      |      |      |
|               | Working Class    | 92              | 13.9 | 90   | 14.1 |             |      |      |      |
| Italy         | Upper Middle     | 69              | 13.2 | 92   | 13.1 | 17          | 11.4 | 21   | 11.8 |
|               | Middle Class     | 58              | 13.3 | 52   | 13.3 | (6th grade) |      |      |      |
|               | Working Class    | 97              | 13.6 | 99   | 13.9 |             |      |      |      |
| United States | Upper Middle     | 112             | 13.5 | 96   | 13.8 | 89          | 13.6 | 77   | 13.7 |
|               | Middle Class     | 290             | 13.6 | 341  | 13.6 | (8th grade) |      |      |      |
|               | Urban Low Income | 94              | 13.9 | 87   | 14.1 |             |      |      |      |
| Austria       | Gymnasium        | 187             | 14.3 | 112  | 14.4 | 29          | 12.4 | 29   | 12.5 |
|               | Hauptschule-A    | 89              | 14.3 | 102  | 14.4 | (6th Grade) |      |      |      |
|               | Hauptschule-B    | 53              | 14.4 | 52   | 14.7 |             |      |      |      |
| Germany       | Gymnasium        | 61              | 14.6 | 60   | 15.0 | 45          | 13.3 | 47   | 13.6 |
|               | Realschule       | 59              | 14.9 | 60   | 15.5 | (7th grade) |      |      |      |
|               | Hauptschule      | 54              | 14.9 | 53   | 15.0 |             |      |      |      |
| Great Britain | Direct Grant     | 39              | 14.3 | 48   | 14.2 |             |      |      |      |
|               | Grammar          | 56              | 14.3 | 61   | 14.3 |             |      |      |      |
|               | Comprehensive    | 87              | 14.4 | 85   | 14.3 |             |      |      |      |

## METHOD OF PRESENTATION

For all groups studied, the tasks were administered by a team consisting of authors and collaborators, with the tasks being presented by a person (E) thoroughly familiar with the language. After being introduced, E explained that the class would participate in a research project intended to show future teachers the wide variety of solution methods used by students to approach certain scientific and mathematical tasks designed for that purpose. The two tasks were next presented as described below, first "Proportional Reasoning" and then "Control of Variables." Finally, students were invited to write on the back of one answer page what they thought about the research project and their participation in it.

We made an effort not to press students for time. After about two-thirds of a class had handed in their answers to the first task, E called for attention and introduced the second task, allowing students who had not completed the first one to keep their papers and finish them later. Approximately fifty minutes allowed most children to complete both tasks, and this period of time either fitted into the ordinary school schedule, into a double period, or could be arranged with help of the school administration. A few students who worked slowly remained during the recess following the testing session in order to complete their papers. Others remained to ask questions about the research project or to use the equipment. In some classes, the time available after all papers were collected permitted class discussion of the tasks, more explanation of the purposes of the research, and/or demonstration of collisions between spheres, as the students requested.

Occasionally a special assembly, fire drill, or other unscheduled event reduced the time available to less than forty minutes and may have resulted in incomplete responses. In such cases, one task was dropped from the schedule or the partially used papers were eliminated from scoring. For this reason, and also because some children came late or had to leave early, there are variations in the number of papers scored compared to the number of subjects in each group listed in Table 1.

Since the two tasks were to be used in five different languages, it was necessary to translate the student pages and to formulate the supplementary oral explanation so it could be presented reproducibly. This preparation was carried out in collaboration with science education research groups in the countries to be visited. We believe that the translations were adequate and did not introduce substantial bias, with one possible exception: the word "target," used in the Control of Variables task to distinguish the stationary sphere from the initially rolling sphere, had corresponding terms in Swedish (malkula) and Italian (bersaglio) but not in German or Danish. The terminology used to paraphrase "target" may have been more difficult for some students to grasp. We shall come back to this matter when we compare the performance of populations in differing countries.

## PROPORTIONAL REASONING TASK

The task for assessing proportional reasoning was similar to the Paper Clips Task--Form B described previously.<sup>7</sup> After making their own measurement, students could use proportional reasoning to transform certain units of measurement. Subjects were given answer pages (see Appendix A) and chains of about nine No. 1 "gem" paper clips. They were then told that Mr. Short on their papers had a friend, Mr. Tall, also drawn on a piece of paper, and

that the two figures, when measured by a row of large round buttons placed side by side, were found to be four buttons and six buttons respectively, from floor to head (the data was written on the board). Then the items on the answer pages were read aloud and the subjects were urged to be complete in their explanations about the height of Mr. Tall and the width of the car. Individuals with questions were referred to the data on the board and were encouraged to use their own ideas. The height of the picture of Mr. Short was found to be between six and seven paper clips by virtually all the subjects.

When papers were handed in, they were briefly scanned by a team member, who asked students with incomplete or obscure answers to clarify their reasoning orally or in writing.

Response categories. In this overview of our findings, we shall report scores on the Proportional Reasoning task in terms of four major categories. As explained previously,<sup>7</sup> four categories are not sufficient to describe all explanations in detail, but they do give a good picture of the students' progress toward proportional reasoning. All papers were separately scored by at least two of the authors, who conferred to reach a consensus concerning the interpretation of an unclear statement in the rare cases of disagreement.

Category I (Intuitive): The explanation does not make use of all the data, or uses the data in a haphazard and illogical way. Examples of responses are:

1. (Mr. T) 10 - I just thought about how many paper clips equal one button and tried to figure out how many paper clips there were from there. Answer 10 pc for Mr. Tall.
2. (Mr. T) 7 - Because there are only 7 paper clips in my chain.
3. (Mr. T) 12 - The way I got that Mr. Tall is 12 inches is I just doubled the six buttons. That is how I got the answer.
4. (Mr. T) 10 - I added 6 and 4 together.
5. (Mr. T) 9 - Because every button equals 1-1/2 paper clips. That means that actually there are six paper clips, but there are nine actually. (Note: did not justify the button/paper clip relation in terms of the data given even when asked orally.)
6. (Mr. T) - I have no buttons to work with so I cannot do your answers.
7. (car) 13 - I guessed.
8. (car) 16 - I think Mr. Tall is 8 paper clips. So I think his car should be  $2 \times 8$  to get 16, because his car should be big, so he can get into it.
9. (car) 14 - Because a paper clip equals a button.
10. (car) 7 - Cause you say, what is half of fourteen.
11. (car) 28 - Because it is 2 paper clips which equals 1 button so I add 14 and 14. (Note the reversal of the button-paper clip relationship.)

Category A (Additive): The explanation focuses on a single difference (tall-short or paper clips-buttons) uncoordinated with other differences, and solves the problem by addition. Examples of responses are:

1. (Mr. T) 8 - If Mr. Tall is 6 buttons and Mr. Short is 4 b, that is a difference of 2. Now Mr. Short is 6 pc tall, so I took the 2 and added it to 6 and got 8.
2. (Mr. T) 8 - Because in buttons he was 2 buttons taller, so I thought paper clips are almost the same length. So I put two paper clips bigger. Because buttons and paper clips should be about the same.
3. (Mr. T) 8 - I got 8 because if Mr. Short is 4 buttons, and is 6 paper clips, then that means Mr. Tall should be 8 paper clips, because he is 6 buttons. The main difference is that there is 2 numbers between the buttons and paper clips.
4. (car) 12 - As in the first example, there are two more clips than buttons, so I subtracted 2 and got 12 buttons.
5. (car) 16 - In buttons it measures 2 more than in paper clips, so if it is 14 in pc's, it will be 16 in buttons. (Note the reversal of the button/paper clip relationship, which occurred in about 5 percent of category A responses.)

Category Tr (Transitional): The explanation shows only partial proportional reasoning, or makes reference to concrete comparisons or iterations. Examples of responses are:

1. (Mr. T)  $7\frac{1}{2}$  - I divided four into six, four is how many buttons Mr. Short is and 6 is the amount of paper clips, and I got  $1\frac{1}{2}$ . Then I added 6, the amount of buttons of Mr. Tall, to  $1\frac{1}{2}$  and got  $7\frac{1}{2}$ .
2. (Mr. T) 9 - First I folded Mr. Short in half, then into fourths, then I measured one of these fourths with my paper clips. Then to find Mr. Tall's height in paper clips I multiplied 6 (from his height in buttons) by  $1\frac{1}{2}$  and got 9.
3. (car)  $9\frac{1}{2}$  - I laid out a line of 14 paper clips on the desk and marked off buttons.
4. (car)  $9\frac{1}{2}$  - If 6 paper clips = 4 buttons, 4 for each time 6 went into 14. It went 2 times evenly, that was 8 buttons, and two clips left. And 2 clips =  $1\frac{1}{2}$  buttons. Added together they equal  $9\frac{1}{2}$  buttons.
5. (car) 21 - Since 6 paper clips = 4 buttons, then 14 paper clips = N buttons.  $\frac{6}{4}$  is  $N/14$ . Cross multiply,  $6 \times 14 = 84 : 4 = 21$ . The answer is 21 buttons. (Note the reversal of the button/paper clip relationship, which occurred in about 15 percent of responses that would otherwise have been classified in Category R.)

**Category R (Ratio):** The explanation uses a proportion or derives the exact scale ratio from the Mr. Tall/Mr. Short dimensions or button/paper clip relationship. No concrete or iterative procedures are employed. Examples of responses are:

1. (Mr. T)  $9 - 1 \text{ button} = 1\text{-}1/2 \text{ paper clips}, 6/4 = 1\text{-}1/2/1, 1\text{-}1/2 \times 6 = 9.$
2. (Mr. T)  $9 - \text{I got this by putting their height in buttons into a fraction } (4/6) \text{ and by putting their height in paper clips into a fraction } (6/x) \text{ and solved it. The result is } 9.$
3. (Mr. T)  $9 - \text{Since } 6 \text{ paper clips} = 4 \text{ buttons, then } R \text{ paper clips} = 6 \text{ buttons, } 6/4 \text{ as } R/6. \text{ Cross multiply } 6 \times 6 = 36, 36 \div 4 = 9. \text{ The answer is } 9 \text{ buttons.}$
4. (Mr. T)  $9 - 6 \text{ clips is equal to } 4 \text{ buttons, } 3 \text{ clips is equal to } 2 \text{ buttons, take half and add. } 6c = 4b, 3c = 2b, 9c = 6b.$
5. (car)  $9\text{-}1/3 - \text{I put his height in paper clips over the width of his car in paper clips } (9/14), \text{ then } (9/14), \text{ then put his height in buttons over his car's width in buttons } (6/x), 6/x = 9/14, 84 = 9x, 9\text{-}1/3 = x.$

#### CONSISTENCY OF RESULTS AND COMPOSITE SCORE ON PROPORTIONAL REASONING TASK

One question we shall answer before proceeding further relates to the consistency of the two parts of the Proportional Reasoning task. Table 2 shows the correlation of scores on the two parts for the more than 3300 boys and girls in the principal groups participating in this study. Each entry is the average percentage of students (calculated from the percentages of fourteen subpopulations of seven countries and two sexes) who responded with the "Mr. Tall" explanation identified at the beginning of the row and the "car" explanation identified at the head of the column. For example, the "14" in location R-Tr of the table means that an average of fourteen percent gave a Category R explanation for "Mr. Tall" and a Category Tr explanation for the "car."

Table 2 Correlation of Scores on the Two Questions of the Proportional Reasoning Task (Percent)

|                              |    | Score on "car" question |    |    |    |
|------------------------------|----|-------------------------|----|----|----|
|                              |    | I                       | A  | Tr | R  |
| Score on "Mr. Tall" question | I  | 13                      | 0  | 1  | 0  |
|                              | A  | 11                      | 14 | 1  | 0  |
|                              | Tr | 3                       | 0  | 11 | 2  |
|                              | R  | 5                       | 0  | 14 | 25 |

Four of the six substantial entries in Table 2 are along the diagonal, representing the same category of explanation on the two parts. A total of sixty-three percent of all papers was of this kind, thirteen percent Intuitive, fourteen percent Additive, eleven percent Transitional, and twenty-five percent Ratio. The other two substantial entries are in the A-I combination and the R-Tr combination, indicating about a third of the students answering in Categories A or R for Mr. Tall's height fell back to the next lower category when applying their reasoning to the car's width. Thus, the car question was more difficult than the Mr. Tall question. Kendall's "tau" rank order correlation coefficient<sup>17</sup> of the two items was 0.48, implying a very good reliability for the task consisting of only two items and offering only four response categories.

The only appreciable differences between boys and girls were in two entries: A-A, which includes seventeen percent for the girls but only ten percent for the boys; and A-I, which includes fourteen percent for the girls and eight percent for the boys. The differences are statistically significant,  $p < 0.01$  by the chi-squared test which we used to evaluate the statistical significance for all comparisons in this report. That is, girls tended to respond in the Additive category more than boys, especially on the Mr. Tall question. The compensating excess of boys was distributed over all the other table entries.

It is gratifying to note that the remaining ten entries in Table 2 are much smaller than the six we have discussed, adding up to only twelve percent of all the responses. For us, this was evidence that Categories I and A do not involve proportional reasoning, while Categories Tr and R do. At the same time, we were reassured that the interaction among students and copying of answers, which certainly took place in many of the classrooms where several students sat at the same table, did not so randomize the responses as to invalidate our method of data gathering.

To simplify the data further, we took advantage of the clustering of responses in Table 2 to create a composite score of I, A, Tr, or R for each student. The composite score assigned to each of the Table 2 locations is indicated in Table 3. If the score on both explanations was the same (diagonal entries), then this was also used as the composite score. If the two were different, then Category R combined with any other category was scored as Tr, Tr combined with A was scored as A, and either Tr or A combined with I was scored as I. Since the student's performance on the two parts was well correlated, the composite score does not mislead by concealing great disparities in the levels of the two explanations.

When the entries in Table 2 are combined as just described, one obtains Table 4, which displays the percent of students having a composite score in each of the four categories. Later in this article we shall present the distribution of composite scores for each of the principal populations listed in Table 1.

#### CONTROL OF VARIABLES TASK

The task for assessing students' reasoning concerning control of variables made use of the collision of two spheres rolling on a track. After identifying the important variables, students could apply the principle that only one variable at a time may be changed if an experiment is to be conclusive.

Table 3 Composite Score on Proportional Reasoning Task

|                              |    | Score on "car" question |    |    |    |
|------------------------------|----|-------------------------|----|----|----|
|                              |    | I                       | A  | Tr | R  |
| Score on "Mr. Tall" question | I  | I                       | I  | I  | Tr |
|                              | A  | I                       | A  | A  | Tr |
|                              | Tr | I                       | A  | Tr | Tr |
|                              | R  | Tr                      | Tr | Tr | R  |

Table 4 Distribution of Composite Scores on the Proportional Reasoning Task (percent)

| Category | I  | A  | Tr | R  |
|----------|----|----|----|----|
| Percent  | 28 | 15 | 32 | 25 |

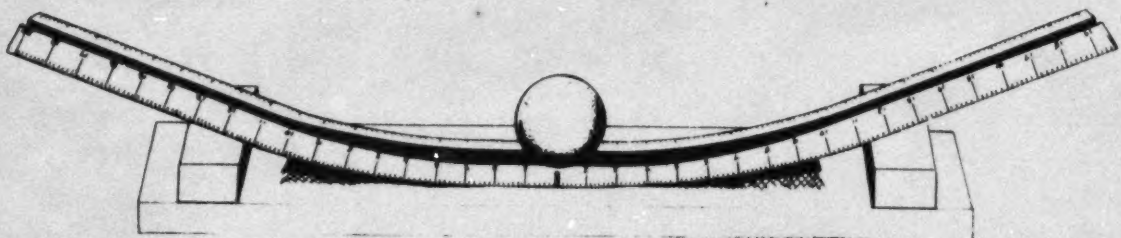


Figure 1. Apparatus for Control of Variables task.

While subjects were previewing the answer pages (see Appendix A), E showed them a track with raised ends (Figure 1), similar to the diagrams on the answer page. One sphere was allowed to roll down and oscillate back and forth. Then a second sphere was placed at the lowest point as "target," and the experiment of allowing a sphere to roll from a "high," "medium," or "low" position to hit the target was described but not performed. E had four equal-sized spheres (two light glass, two heavy steel) which he showed to the class before reading Question 1 aloud and inviting the students to write their answers in the form of a list. After about a minute,

several volunteers were invited to read items from their lists. Then E called attention to the weight of the sphere--an item that was always mentioned--to direct attention to Questions 2 and 3. After these had been read aloud, the students were asked to answer them completely. Questions about the test were invited and answered individually.

Response categories. Here also we shall report our findings in simplified form, and group together responses that appeared to us logically similar though they are qualitatively different.<sup>15</sup> Since Questions 2a and 2b are similar, we adopted the same three-level classification system for both. We assigned zero, one, or two points to a response according to its level, as described below. For Question 3 we distinguished acceptable from unacceptable answers, assigning one point to the former and zero to the latter, as is also described in detail with illustrative examples below. All papers were read by two or more authors, with disagreements regarding interpretation being resolved through discussion.

Questions 2a and 2b:

0 points - any starting position lacking an explanation or accompanied by an "explanation" that describes the experimental set-up, what the subject wishes to find out, or the phenomenon that will occur. Examples are:

1. 2(a) High, high - Well it seems that if you start it high they'll both get a lot of speed which would be better.
2. 2(a) Low, high - If, when the target was hit by the heavy sphere, it went as far as when hit with the light sphere, you can tell that the spheres can have the same effect even though they are of different weights, as long as they are started at places that will give them different rates of acceleration and speed.
3. 2(a) Medium, high - If we start the light sphere at a high spot then it will probably have as much speed and force as the heavy sphere starting at the medium spot.
4. 2(a) Medium, high - I would start the heavy sphere at medium to see if even though the sphere is heavy, will it make a difference? I started the light one at high so it would pick up speed and knock the ball far.
5. 2(a) High, low - When you start the light sphere low, then when you started the heavy sphere high you could see the big difference clearly.
6. 2(b) High - To see how far the target moves
7. 2(b) Medium - The heavy ball will make it fall off.
8. 2(b) Low - To get an accurate amount of height.
9. 2(b) High - It will build up speed and hit the target much further than the light sphere from a low position.

1 point - Equal starting positions accompanied by an "explanation" that ascribes importance, value, or usefulness to this equality.

Examples are:

1. 2(a) High, high - Start them at the same place and give them the same speed, then measure how far the target goes up the other side.
2. 2(a) High, high - If you want to know the difference the weight of the sphere makes, you should start them in the same place.
3. 2(a) High, high - I would start them at the same place, because this would tell me which one is lighter and which is heavier. The one that hits with a greater force is heavier.
4. 2(b) Low - It would be better because we could tell how much of a difference it makes.
5. 2(b) Low - So they'll be equal.

2 points - Equal starting positions accompanied by an "explanation" stating the necessity of choosing this equality and possibly mentioning the principle that only one condition should be varied at one time.

Examples are:

1. 2(a) Any equal - The main reason of this experiment is the weight difference, so you would have to keep all other factors the same.
2. 2(a) High, high - I'd start them both high because that would provide more of a margin of difference to study. They'd have to start at the same point to make the experiment constant because you should have only one variable.
3. 2(a) Medium, medium - You'd have to start both spheres at the same position, otherwise, one would be at a disadvantage. One might gain more speed.
4. 2(b) Low - Low because a good experiment has only two variables, in this case, weight of sphere and how far target goes.
5. 2(b) Low - Because you have to keep all things other than mass constant, as explained above.

Question 3:

0 points - "yes" accompanied by any explanation, or "no" accompanied by an "explanation" that does not refer to either the difference in starting positions or target weights. Examples are:

1. Yes - I figure if the target went farther, then the metal A ball must be better.
2. Yes - Because it is farther up and its target is lighter than B's.

3. Yes - Because metal A isn't going to hit in the same place every time.
4. Yes - Because A is more powerful and heavier.
5. No - Because both of them weigh the same.
6. No - It wasn't a controlled experiment because you really couldn't tell from the information given.

1 point - "No" accompanied by an explanation referring to the starting positions and/or the target weights. Examples are:

1. No - This is not a good proof because one target is light and one is heavy. Metals A and B are set at different points, which would make a difference. The target should be the same weight and the metals A and B should be let go at the same point.
2. No - Metal B was released in the medium position which doesn't allow it as much room to speed up as metal A, so Metal B doesn't have as much force on the ball so it doesn't go as far.
3. No - It doesn't prove it because the targets are different.
4. No - Metal A was at a higher position and hit a lighter target.

#### CONSISTENCY OF RESULTS AND COMPOSITE SCORES ON THE CONTROL OF VARIABLES TASK

We have represented the correlation of the scores on the three questions in Table 5, which consists of two parts. To the left is the correlation of the response-levels to questions 2a and 2b for subjects who scored zero on Item 3, and to the right is the same correlation for subjects who scored one point on Item 3. The entries are average percentages like those in Table 2. On the average, one-third of a population scored zero on Item 3, and almost all of these, twenty-six of thirty-three percent, also scored zero on Items 2a and 2b.

Two-thirds of the subjects scored one point on Item 3. One-third of these, or twenty-two percent of the entire sample, scored zero on both parts of Item 2, which was clearly much more difficult than Item 3. Of those students who did not score zero on Item 2, about half scored differently--usually better on (b) than on (a). Boys and girls performed substantially the same.

Even though there is a good correlation on the two parts of Item 2 (seventy-three percent of the students have the same score), this is brought about by the great difficulty of the item, which has the result that almost half of all the students score zero on both parts. We conclude that the correlation is deceptive, and that control of variables on Item 2a represents a greater accomplishment than control of variables on Item 2b. Kendall's coefficient of concordance<sup>19</sup> for the three questions has the value 0.49, an adequate indication of reliability in view of the fact that the questions had only two or three response categories.

To create a composite score for the Control of Variables task, we have added the number of points each student received on his three explanations. This procedure implies that a more advanced answer on one part can compensate for a less advanced response on another. Even though that is not strictly true since the parts are of differing difficulty, the total number of points does give a good indication of relative performance, since each total score is achieved primarily through one particular point combination on the parts. The distribution of the total number of points is given in Table 6, where we have also indicated the most frequent point combinations on Items 2a, 2b, and 3 for achieving each score. Thus, most students whose composite score was three had one point on each of the three items. Note the small percentage of students having a composite score of four points. We believe it includes primarily subjects who were capable of scoring five but failed to do so out of carelessness or misunderstanding.

Table 5 Correlation of Scores on the Three Items of the Control of Variables Task (percent)

| Score 0 on Item 3 |   |         |   |   | Score 1 on Item 3 |                                     |         |   |    |
|-------------------|---|---------|---|---|-------------------|-------------------------------------|---------|---|----|
|                   |   | Item 2a |   |   |                   |                                     | Item 2a |   |    |
|                   |   | 0       | 1 | 2 |                   |                                     | 0       | 1 | 2  |
| Item 2b           | 0 | 26      | 0 | 0 | Item 2b           | 0                                   | 22      | 1 | 0  |
|                   | 1 | 3       | 1 | 0 |                   | 1 <td>10</td> <td>9</td> <td>2</td> | 10      | 9 | 2  |
|                   | 2 | 2       | 0 | 1 |                   | 2 <td>5</td> <td>3</td> <td>14</td> | 5       | 3 | 14 |

Table 6 Distribution of Composite Scores on the Control of Variables Task (percent)

| Composite score                             | 0     | 1     | 2     | 3     | 4 | 5     |
|---------------------------------------------|-------|-------|-------|-------|---|-------|
| Percent                                     | 26    | 25    | 14    | 14    | 6 | 14    |
| Most frequent pattern of scores on #2a,2b;3 | 0,0;0 | 0,0;1 | 0,1;1 | 1,1;1 | - | 2,2;1 |

# RESULTS FOR DENMARK

All classes in Danish schools are coeducational, and there is no ability grouping in the first seven grades. Beginning with the eighth grade, however, there is a division into two tracks, one that will finish school after the tenth grade and another that enters gymnasium and ultimately completes twelve years of school. Our subjects came from Frederiksberg, a city contiguous with Copenhagen, and three suburban communities within commuting distance of Copenhagen. We met with seventeen classes of seventh graders and two classes of eighth graders, one from each of the two tracks, in early May, 1974. The teachers in these classes specialize in subject areas, and the set of teachers usually remains with one class for several years. Even though there were substantial fluctuations in student performance on the two tasks from class to class, the results for any community or for any one grade were not appreciably different from those for any other. We are therefore reporting only one set of data for Denmark. The frequencies of the various composite categories of responses to the Proportional Reasoning task are given in Figure 2, and those for the Control of Variables task are given in Figure 3. (The significance of the shading will be explained later with the help of Table 7.)

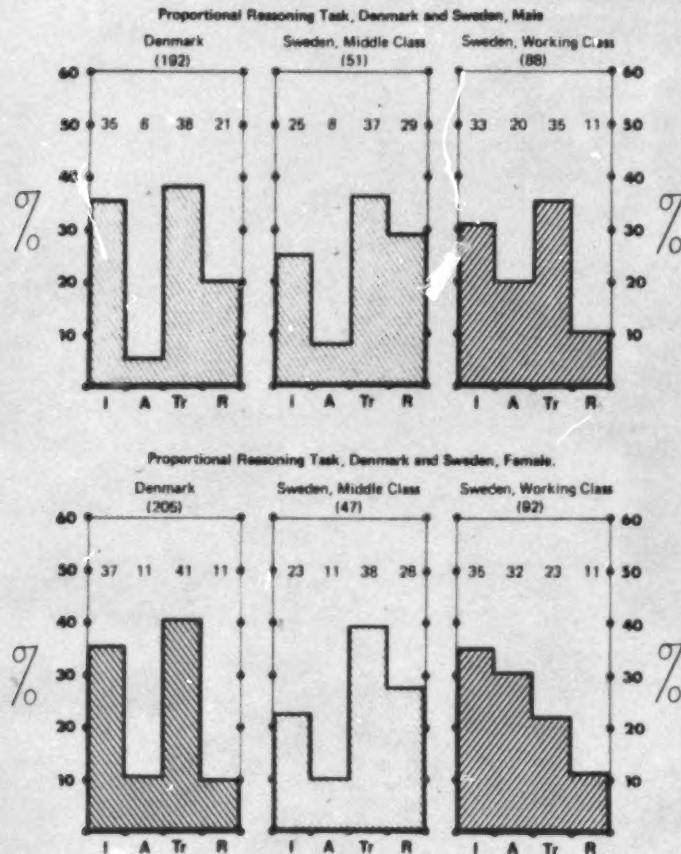


Figure 2. Proportional reasoning data, Denmark and Sweden.

The distributions on both tasks are similar for boys and for girls. On proportional reasoning, somewhat more than one-third were Intuitive and about forty percent were Transitional. In the Ratio category, boys were substantially more numerous, with the difference statistically significant ( $p < 0.15$ ). On control of variables, the differences between boys and girls were small, statistically insignificant. About fifteen percent scored in the two top categories (4 or 5 points), and more than half scored in the two lowest categories (0 or 1 point).

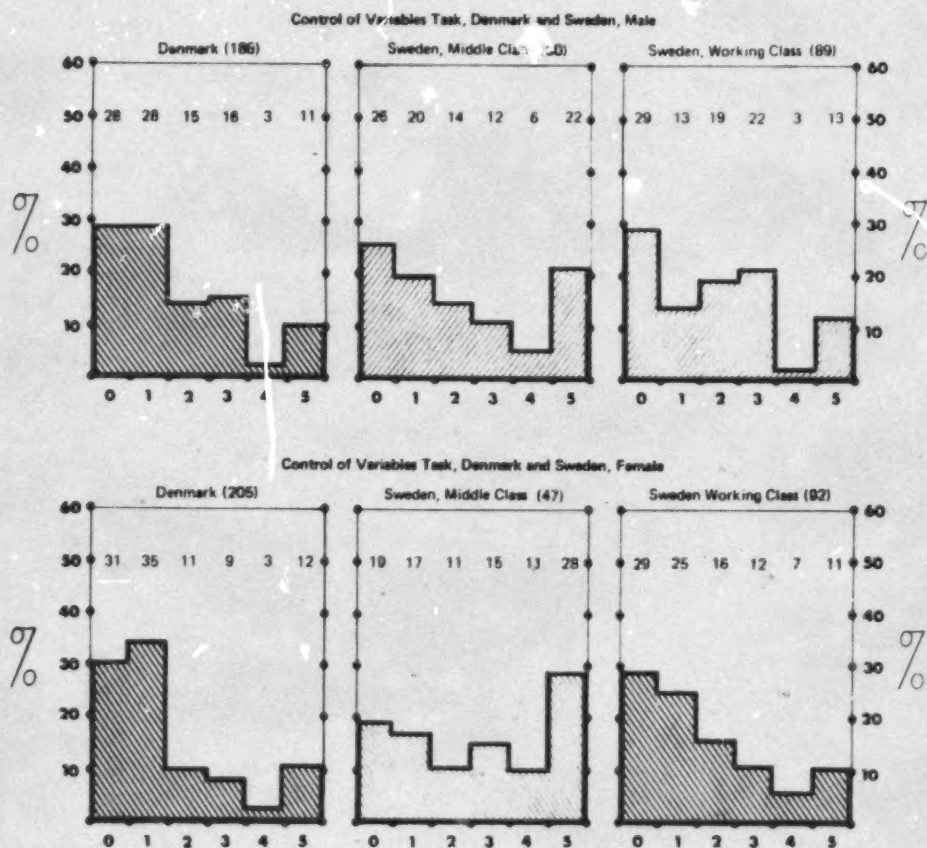


Figure 3. Control of variables data, Denmark and Sweden.

To display the variation of student reasoning from class to class, we present in Figure 4 a scattergram showing all nineteen Danish classes we visited. The abscissa on the graph indicates the percentage of students who answered the Proportional Reasoning task in Categories Tr or R, while the ordinate indicates the percentage of students who scored three or more points on the Control of Variables task. In our judgment, these percentages represent the fraction of students who are at least transitional to formal thought.

## Denmark

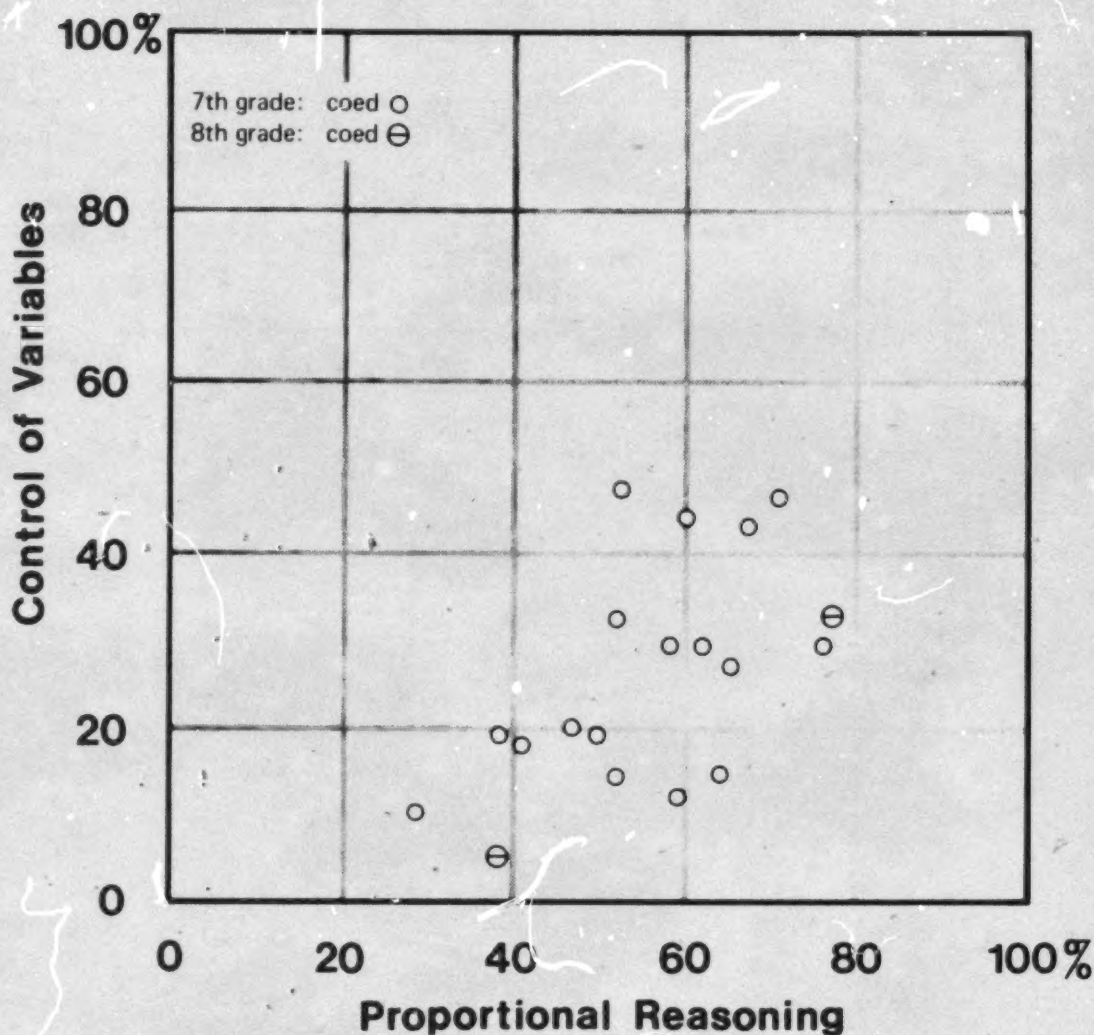


Figure 4., Scattergram of the percentage of Danish students in a class who showed transitional or formal reasoning on each task.

Grouping students together as a class shows the actual teaching situation faced by the science or mathematics instructor of that class. There is quite a difference between the poorest-performing classes near the lower left on the graph, and the best-performing ones near the center. Actually, the low classes were in schools that also contributed classes with very high averages. The two eighth-grade classes clearly show the effect of tracking, though their average is like that of the seventh graders.

## RESULTS FOR SWEDEN

In the coeducational Swedish school, instruction in the seventh and eighth grades is carried out by specialists. The students have several hours of very traditional science instruction in physics, chemistry, and biology each week. Our school visits in Sweden were carried out in Gothenburg and its surroundings during late May, 1974. Gothenburg is an industrial city, with large ship-building, port, and automobile manufacturing facilities. The labor shortage in Sweden has led to a substantial immigration of Finnish workers whose children, speaking a different language, have some difficulties adjusting to the local schools.

Our student sample was taken from (1) two working class neighborhoods consisting of fairly large modern apartment houses and (2) a suburban community with large and small apartment houses. We have designated these as working class and middle class populations, respectively. The working class children were all seventh graders, while the middle class children were half seventh graders and half eighth graders. Because of the small size of the middle class sample and the virtually identical scores for the two grades, we combined the data for the two grades and are reporting on a population of greater average age (Table 1).

The frequencies of the various categories of responses on the Proportional Reasoning and Control of Variables tasks are given in Figures 2 and 3 beside the Danish results. In proportional reasoning, almost thirty percent of middle class students scored in the Ratio category, but only ten percent of the working class students did so, a statistically significant difference ( $p < 0.01$ ). On the Control of Variables task, the differences also favored the middle class group, but were smaller. The differences between boys and girls of the same socioeconomic group were slight.

To display the variation of student reasoning from class to class, we present in Figure 5 a scattergram similar to that in Figure 4 for Denmark. This diagram shows the group differences very clearly. There were two very successful middle class groups, one seventh and one eighth grade, which had been pointed out to us as unusual on the basis of test scores and teacher opinion. The other middle class seventh- and eighth-grade groups, and all eight seventh-grade working class groups, cluster near one another in the lower portion of the graph. The effect of the increased age and school experience of the eighth graders is seen to be small, as it was in Denmark.

## Sweden

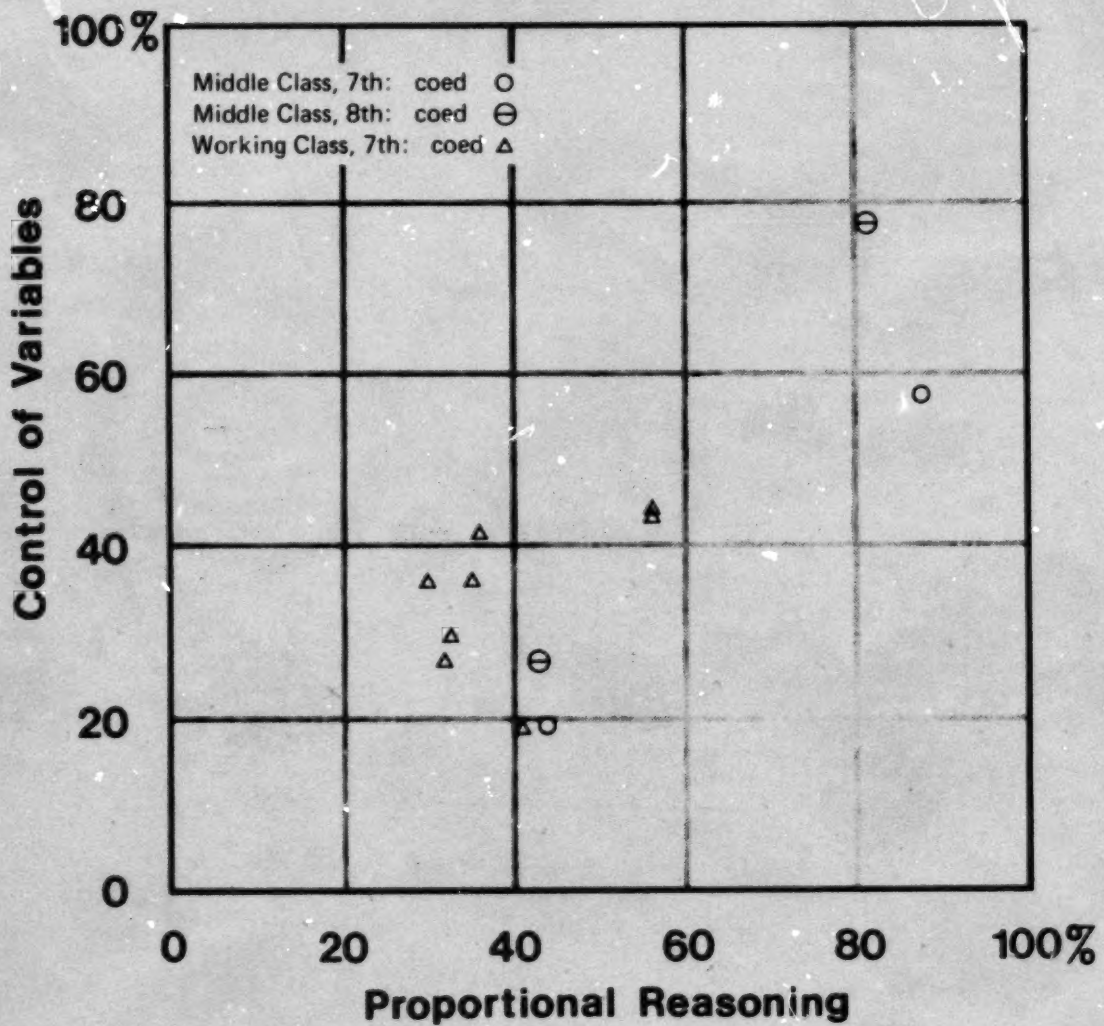


Figure 5. Scattergram of the percentage of Swedish students in a class who showed transitional or formal reasoning on each task.

# RESULTS FOR ITALY<sup>20</sup>

Italian eighth graders are in the last year of the "media," a three-year middle school. The media is a comprehensive school that may be co-educational or not. The various subjects are taught by academic specialists who usually remain with a group of students for the entire three years. After the media, some children discontinue school while others attend secondary schools for technical or vocational studies or university preparation.

The schools participating in our study in February, 1974 were located in Rome and were chosen to sample children from various socioeconomic levels. Boys and girls of the working class attended the same schools, but studied in classroom groups that were segregated by sex. Some middle class children attended coeducational classes while others were segregated. The upper middle class children came from a coeducational school and from two specialized schools, a private university preparatory school for boys and a school for girls featuring dance and dramatic arts.

The frequencies of the various response categories on the Proportional Reasoning and Control of Variables tasks are given in Figures 6 and 7. There are many differences among the groups, and we shall point out the most noteworthy of these.

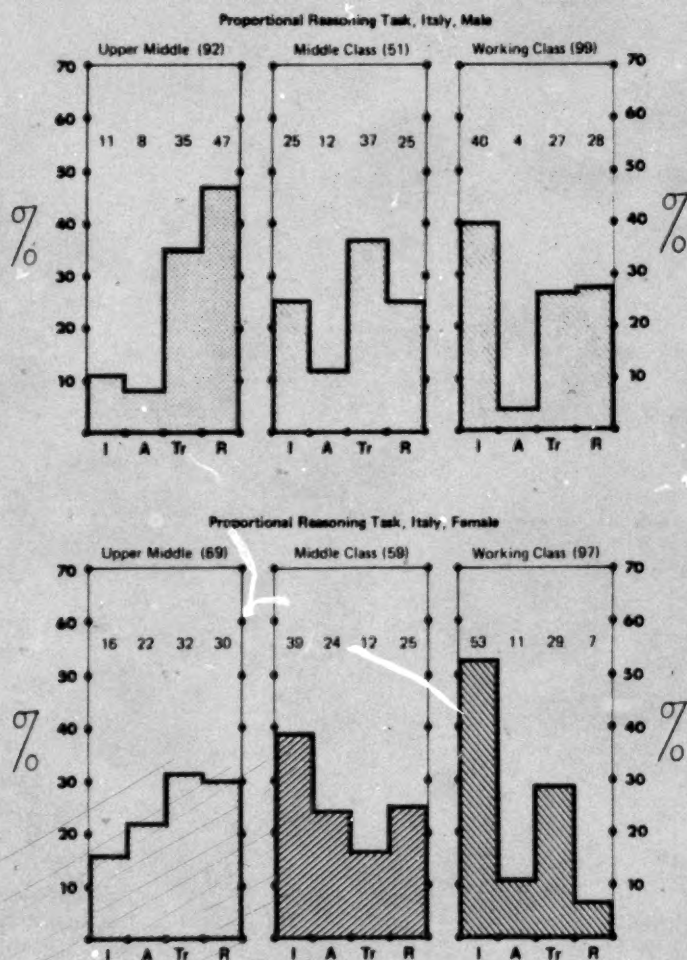


Figure 6. Proportional reasoning data, Italy.

On the first task, the boys at all socioeconomic levels showed more proportional reasoning than the girls. The fractions of Transitional and Ratio responses of upper middle class, middle class, and working class boys were 82 percent, 62 percent, and 55 percent, respectively, while the same fractions for the girls were 62 percent, 37 percent, and 36 percent. It should be added, however, that the differences were much smaller than this in the coeducational classes that are part of our sample. The same figures reveal a substantial effect of socioeconomic level on performance. Girls responded much more frequently than boys with an Additive response; for instance, 17 percent of all Italian girls gave Additive answers to both questions on the Proportional Reasoning task, and another 26 percent gave at least one Additive answer. Of the boys, only 5 percent gave both Additive answers and 6 percent gave one. Boys were slightly higher than girls on Intuitive responses (20 percent to 14 percent, statistically significant only at the five-percent level).

On the Control of Variables task, the differences between boys and girls were much smaller and are not significant statistically. The differences between the middle and upper middle class groups were also quite small, but the working class children found the task very difficult and more than 70 percent scored one point or less. The other groups found the task difficult also, but not quite so difficult, with 40 to 55 percent scoring in the two lowest categories.

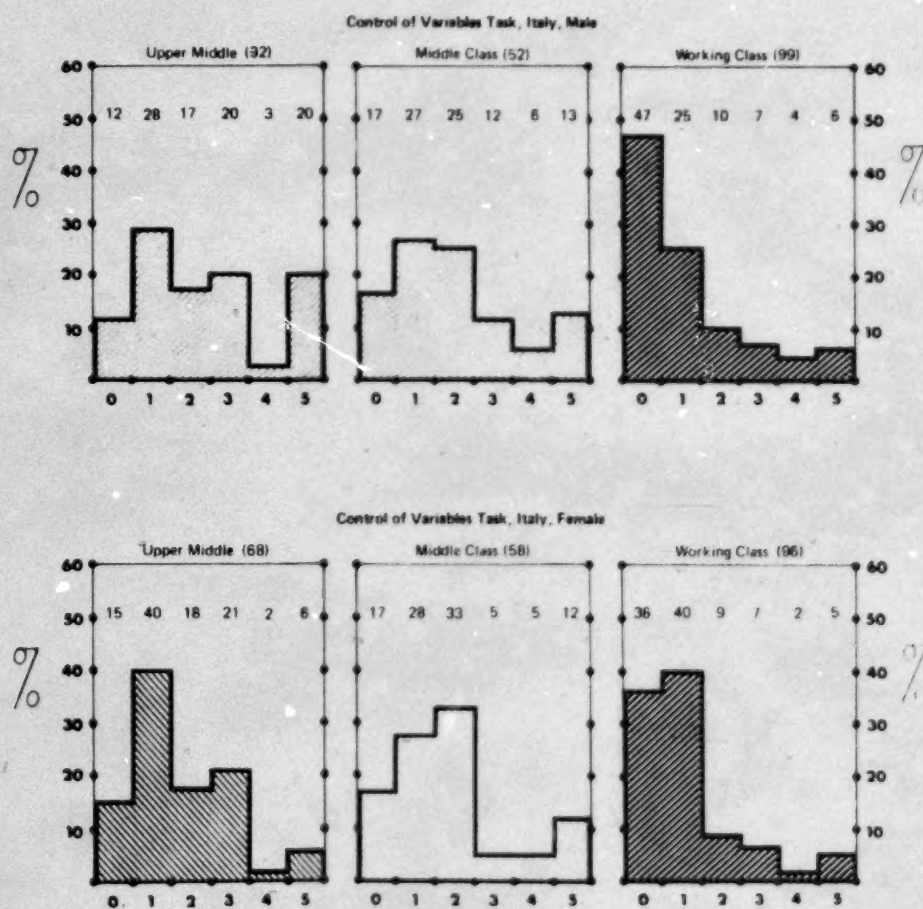


Figure 7. Control of variables data, Italy.

A scattergram of the class to class variation of the Italian students is presented in Figure 8. The overall pattern of points is below the diagonal, again indicating that the Control of Variables task was more difficult than proportional reasoning for the Italian-children. The socioeconomic influence is visible, but there is a good deal of overlap among the three levels. One cluster of six points to the upper right of the diagram includes the three coeducational upper middle class groups, the two upper middle class boys groups, and one working class group of boys taught by an outstanding mathematics teacher in a school known for its innovative practices. At the other extreme is a cluster of three points, two classes of girls and one of boys from working class neighborhoods. Near the middle of the diagram you can find all the other classes, some from each socioeconomic level. Here also are two points representing two upper middle class coeducational pilot groups of sixth graders (first media) attending the same school as the eighth graders. There is a substantial difference between the two ages.

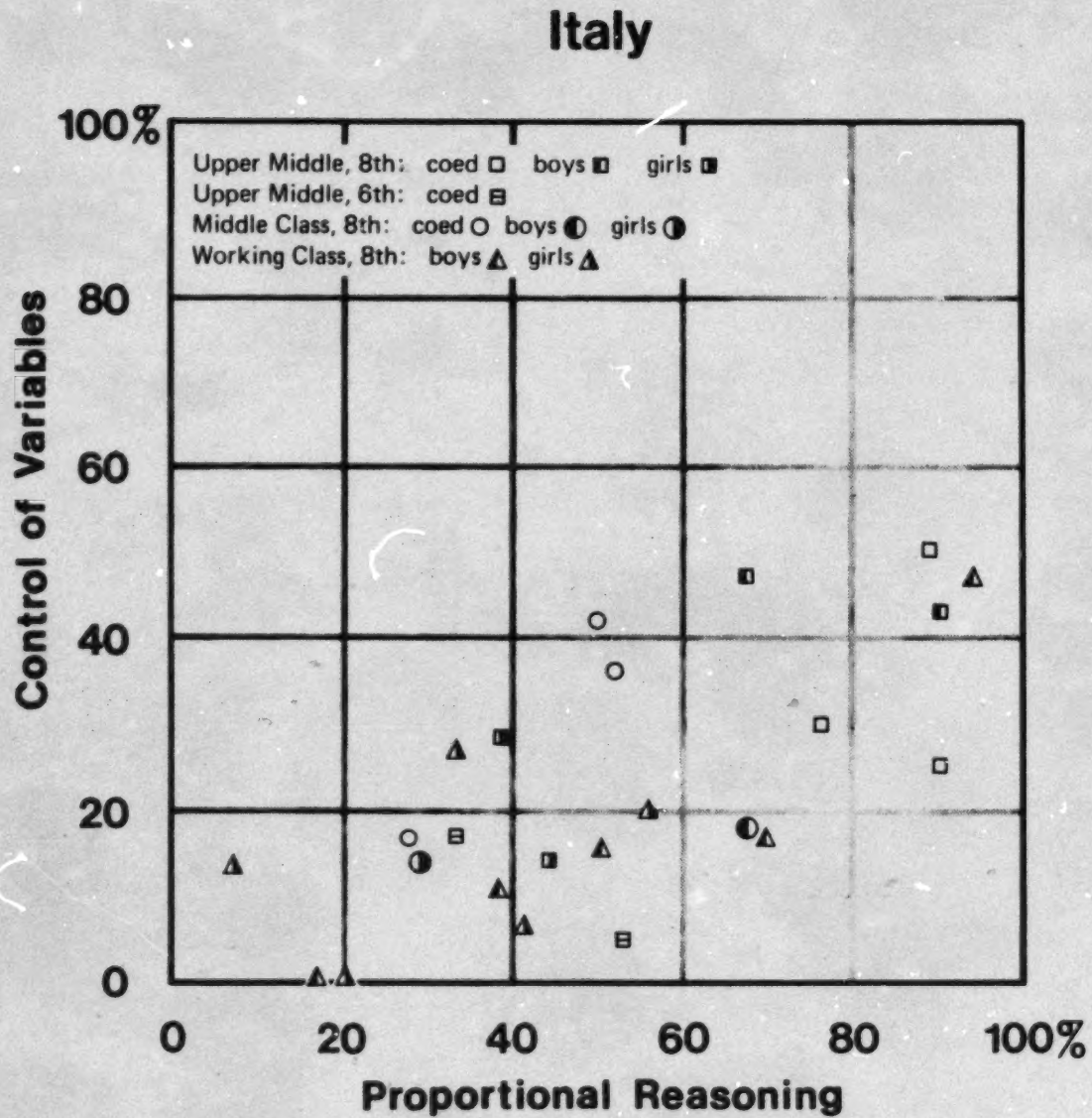


Figure 8. Scattergram of the percentage of Italian students in a class who showed transitional or formal reasoning on each task.

# RESULTS FOR THE UNITED STATES

The subjects in the United States came from nine different cities and towns in the northeastern and north-central part of the country. Their socioeconomic levels were determined from the character of the community in which they lived. We identified three levels, upper middle class suburban, middle class urban and suburban, and urban low income neighborhoods. All schools were coeducational, some covered grades 7-8, some 7-9, some 1-8, and some 7-12. Many practiced a form of tracking by academic achievement or foreign language preference. By selecting classes from several ability levels, we hoped to reduce the sample biases associated with tracking. The teachers were subject matter specialists, but there was no continuity of class groupings or teachers faced by a particular student from year to year. Our investigation occurred between October, 1973 and January, 1974.

The frequencies of the various categories of responses on the Proportional Reasoning and Control of Variables tasks are given in Figures 9 and 10. As in Sweden and Italy, formal thought as indicated by the two tasks is statistically related to socioeconomic status. About forty percent of middle and

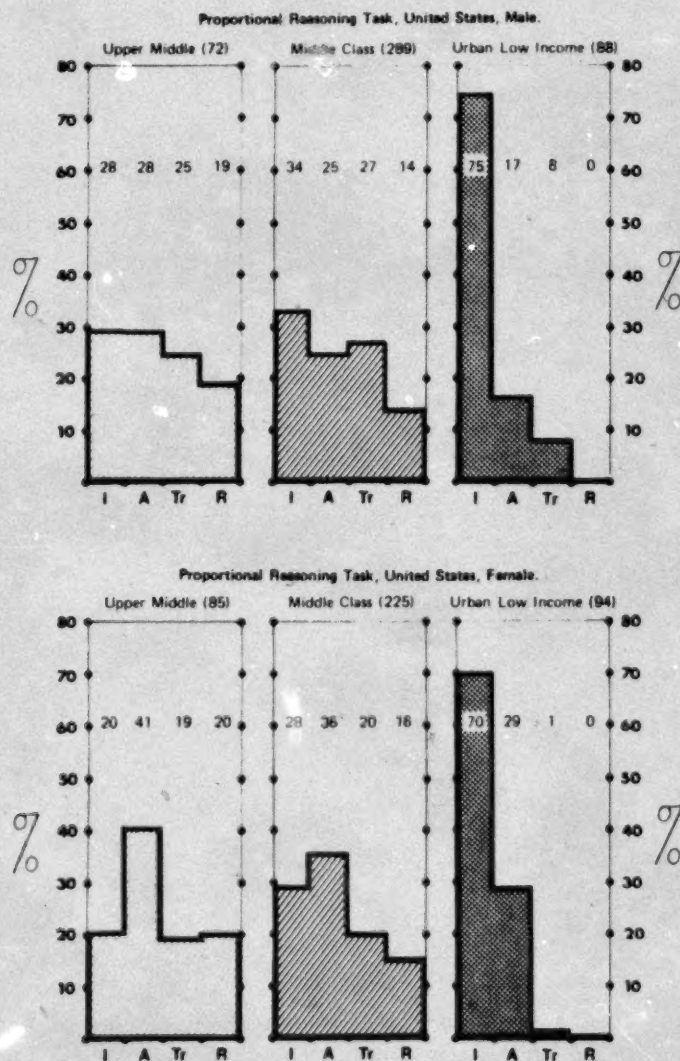


Figure 9. Proportional reasoning data, United States.

upper middle class students responded in the Transitional or Ratio categories of the Proportional Reasoning task, while fewer than ten percent of the urban low income youngsters did so. Boys and girls scored very similarly, with the girls giving more answers in the Additive category at all three socioeconomic levels, while the boys responded more frequently in the Intuitive and Transitional categories. These small differences were statistically significant only for the very large middle class group.

On the Control of Variables task, the urban low income boys scored significantly better than the girls ( $p < .01$ ): forty-six percent of the boys earned one or more point, but only nineteen percent of the girls did so. Differences between boys and girls in the other socioeconomic levels were very slight. Somewhat more than fifty percent of the upper middle class children scored three or above, compared to about thirty-five percent for the middle class children.

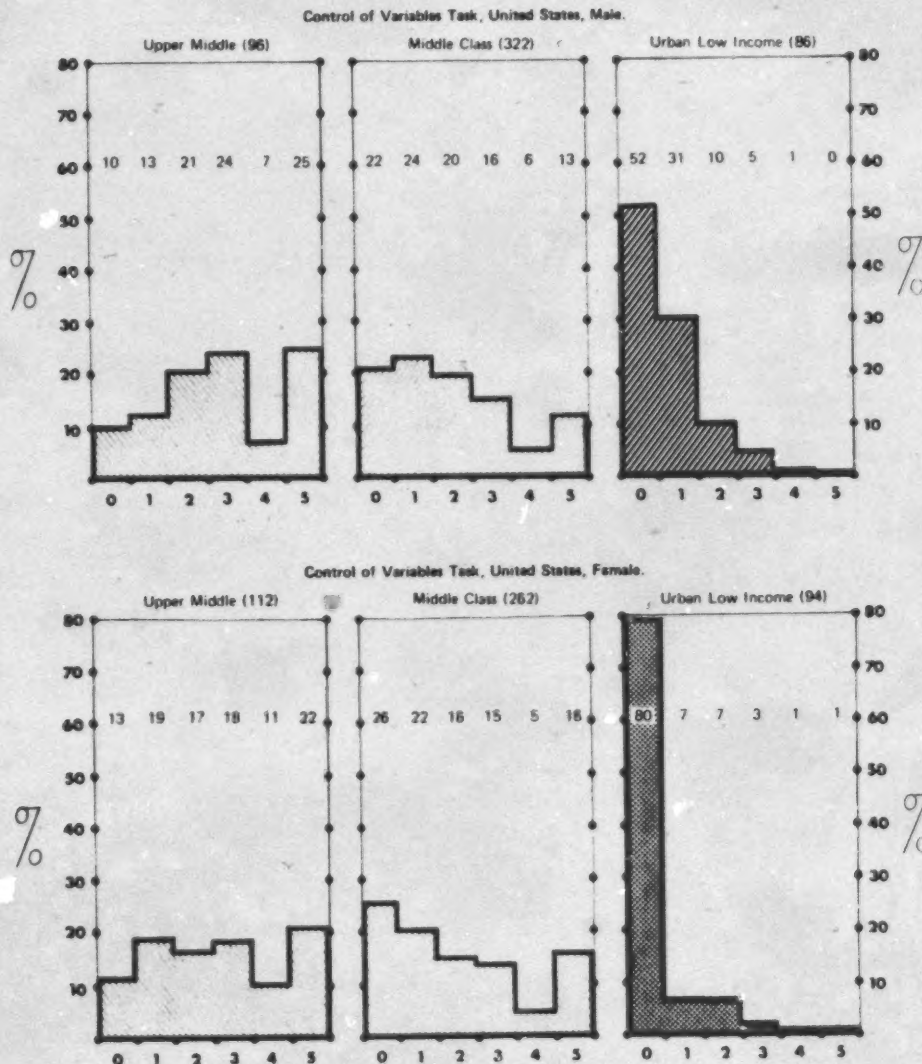


Figure 10. Control of variables data, United States.

## United States

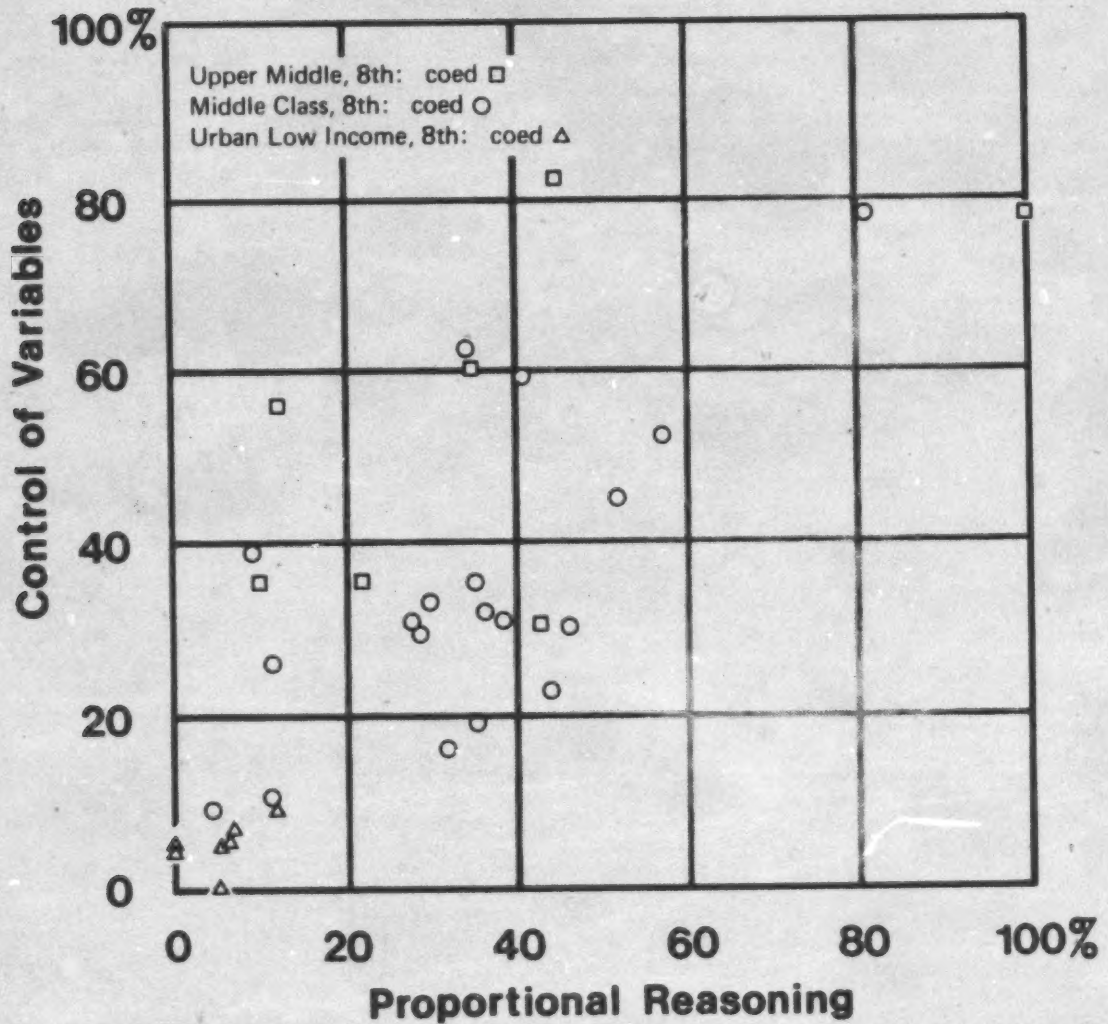


Figure 11. Scattergram of the percentage of United States students in a class who showed transitional or formal reasoning on each task.

The variation in explanations among classroom groups is shown in Figure 11, a scattergram similar to Figure 4. One can see that the points are distributed generally above the diagonal, indicating that the Proportional Reasoning task was the more difficult one for the students in the United States. The clustering of urban low income classes in the lower left corner of the diagram reveals once again these students' great difficulty in formulating a successful explanation; many had left the answer spaces on their papers blank and did not give an oral statement of their ideas to the team of investigators. Located with these classes in the diagram there are also two middle class groups of children with diagnosed learning difficulties.

Toward the upper part on the diagram are mostly groups from the upper middle class suburban communities, but they do not fall further to the right (proportional reasoning) than many middle class groups. In only two classes, both accelerated, did substantially more than half the students show proportional reasoning in the Transitional or Ratio categories.

#### RESULTS FOR AUSTRIA

Austrian students attend secondary school beginning with the fifth grade. Two school types are available, the Gymnasium offering an eight-year program leading to university entrance, and the Hauptschule offering a four-year program that may lead to a vocational school, a technical school, or apprenticeship and employment. The Hauptschule itself is divided into two academic tracks designated as Track A (more able students) and Track B (less able students). Many Gymnasiums are presently for boys or girls only, but there is a trend toward coeducation. The Hauptschule is basically a coeducational, departmentalized middle school.

In the city of Vienna, the location of all the schools we visited in March of 1974, both school types are easily accessible to all students, with the result that parents can send their fifth graders who meet the academic requirements to any Gymnasium in the city. About half the students in any age group in Vienna attend Gymnasium, about thirty percent are in Track A, and about twenty percent are in Track B. (In rural areas there are fewer Gymnasiums; therefore a larger percentage of students attends Hauptschule outside Vienna.) All teachers are academic specialists. Even though the two school types are administratively separated, there is an effort to coordinate the curricula so that students from Track A of the Hauptschule can transfer to the Gymnasium after a year or two of special effort.

The frequencies of the various categories of responses on the Proportional Reasoning and Control of Variables tasks are given in Figures 12 and 13. On the Proportional Reasoning task the Gymnasium students performed outstandingly, with more than eighty percent of boys and girls showing transitional or proportional reasoning. Sex made very little difference here, even though most students tested were in segregated schools. A frequent form of the explanation for Mr. Tall's height and the car's width appealed to the "rule of three," a procedure for solving problems of proportions that is taught in the Austrian sixth grade. Most of the students who used this method applied it correctly. Another frequently used procedure took advantage of the rulers which virtually all students brought with them. After measuring Mr. Short's height and the paper clip length in centimeters, these students calculated the size of a button and the height of Mr. Tall in centimeters, and finally his height in paper clips. We scored such an answer in the Transitional category since it seemed to us similar to the other concrete procedures for dividing up Mr. Short.

It is interesting to see from the data that the Track A (Hauptschule) boys scored almost as well as the Gymnasium students, but that the Track A girls, with only fifty percent Transitional or Ratio, scored very substantially lower ( $p < .001$ ). The Track B boys' responses very closely followed the pattern of responses of the Track A girls, while fewer than thirty percent of the

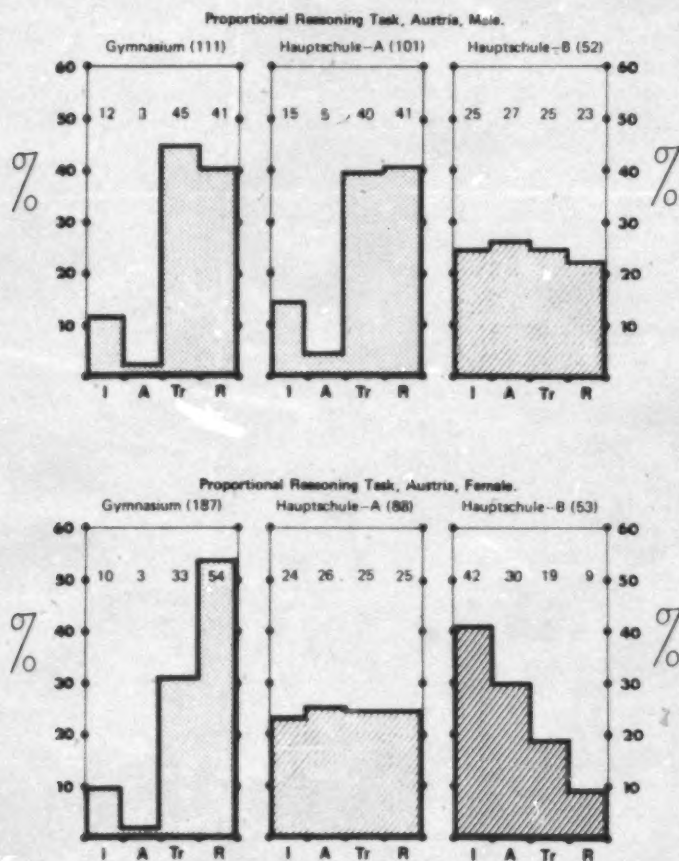


Figure 12. Proportional reasoning data, Austria.

Track B girls scored in Categories Tr or R. These substantial differences between boys and girls in the same tracks and in the same classrooms are quite surprising. They suggest that the criteria according to which the more gifted students attend the Gymnasium may operate very differently for boys and girls, at least insofar as proportional reasoning is concerned.

On the Control of Variables task, the performance of boys and girls in the same courses was much more similar, with the largest difference in favor of the boys occurring for the Gymnasium students ( $p < 0.05$ ) and the smallest for the Track B students. The Track B students had very great difficulty--fewer than five percent scored more than one point. The differences between academic levels were significant ( $p < 0.005$ ) except that between the Track A and Track B girls. The students' questions during the administration of this task indicated that the wording was difficult for many to understand. Especially Item 2b, which had elsewhere been answered more easily than 2a, was left to the end by many students who asked for an additional oral explanation and encouragement before proceeding.

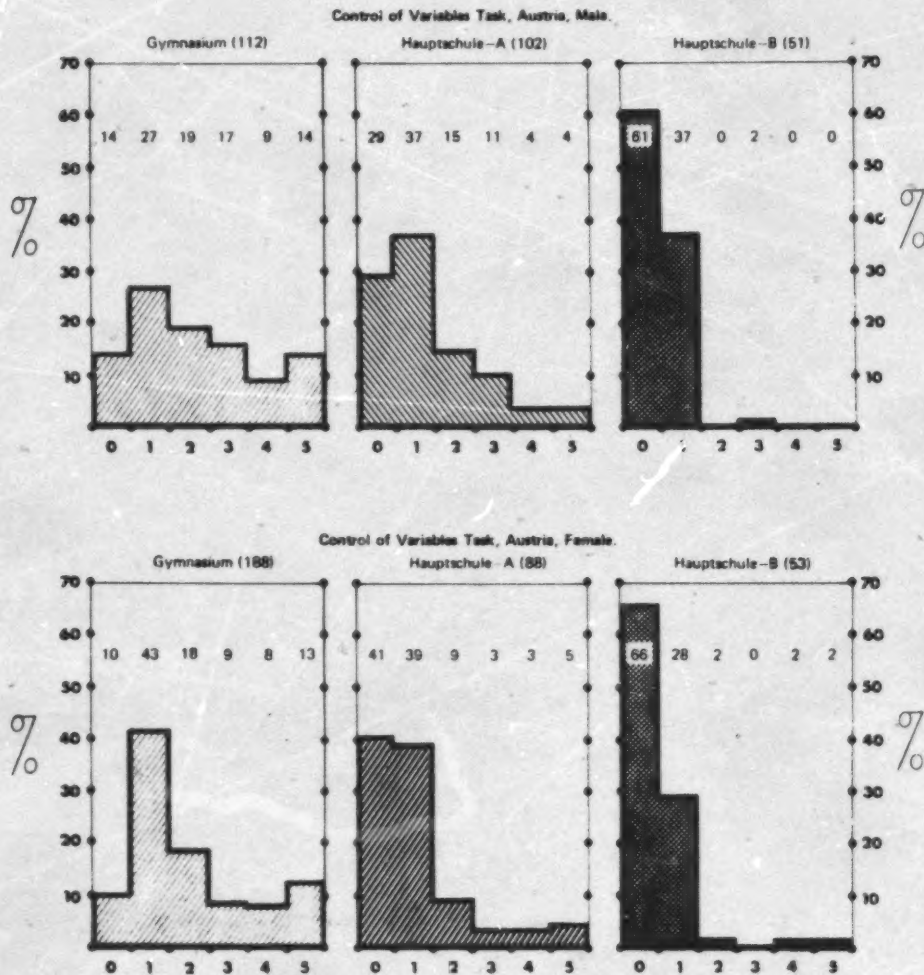


Figure 13. Control of variables data, Austria.

The scattergram in Figure 14 shows the variation in student reasoning from class to class. The clustering of all points below the diagonal of the graph reflects the children's difficulty with the Control of Variables task. The many Gymnasium points along the right side lend emphasis to the children's outstanding success on the Proportional Reasoning task. The differences between the three tracks can again be seen clearly in the grouping of the points in the diagram. The boy-girl differences are concealed, however, and tend to reduce the contrast between Tracks A and B of Hauptschule.

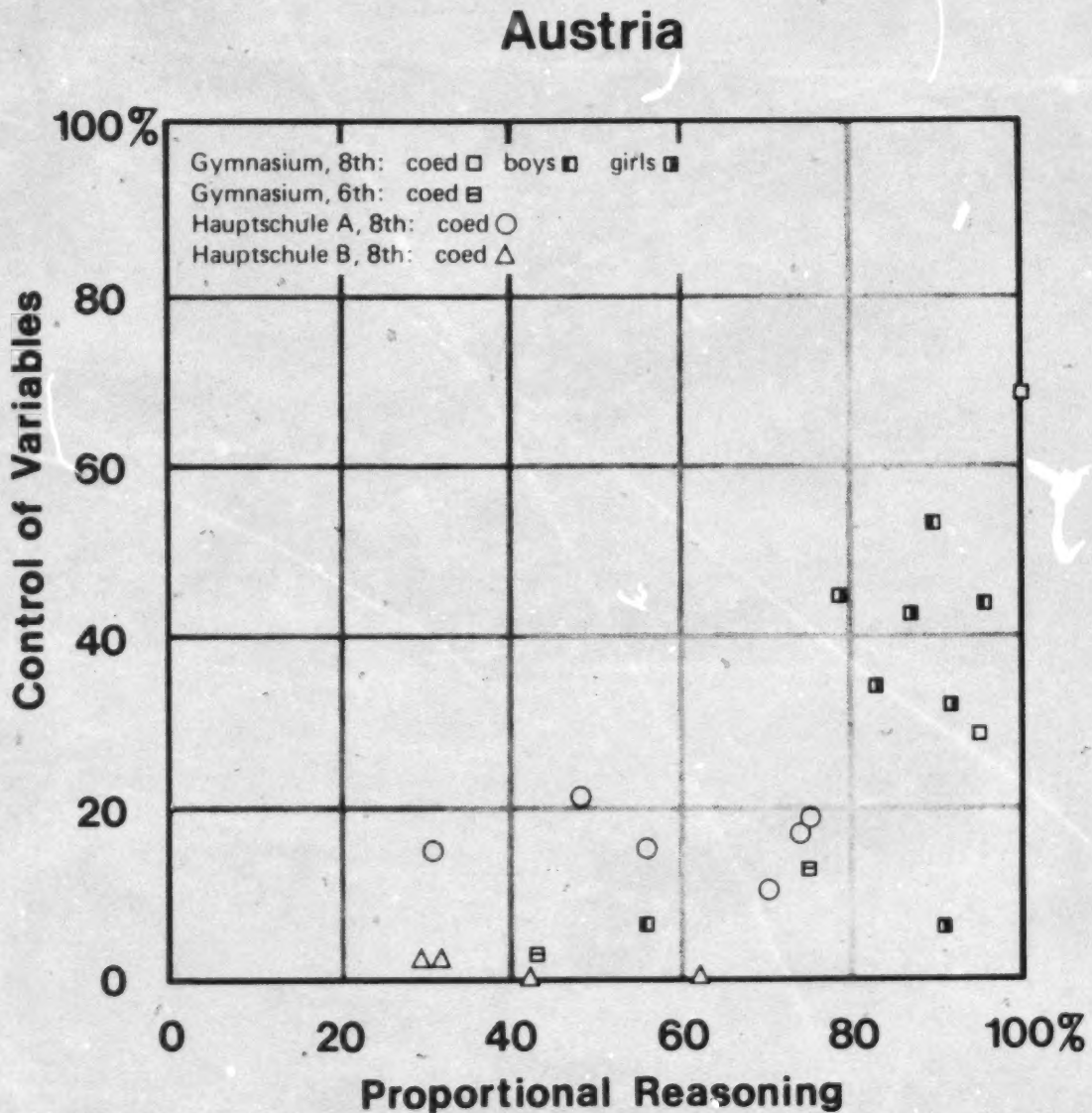


Figure 14. Scattergram of the percentage of Austrian students in a class who showed transitional or formal reasoning on each task.

Note that there is one Gymnasium class that falls in the group of Track A classes, near the lower middle of the graph. We were not able to identify any special circumstance that could explain this observation. Note also the two points of sixth grade Gymnasium classes (black circles) that were tested as pilot groups. Their performance was substantially weaker than that of their older schoolmates, but close to that of the two-year-older Hauptschule students. They clearly have not yet mastered the Proportional Reasoning task.

#### RESULTS FOR GERMANY

The German school system presently offers one coeducational, comprehensive neighborhood elementary school for all children in the first four grades. The secondary schooling beginning with the fifth grade consists of three alternative types, with selection for the more challenging courses being based on grades and teacher recommendations. (An optional entrance examination is offered for children whose parents wish to contest the assignment of their child). Most demanding is the eight-year Gymnasium, which prepares students for university entrance and enrolls about twenty percent of an age group. Next is the Realschule, which offers a six-year course leading to tertiary education in a technical school or specialized academy, and which enrolls about twenty-five percent of an age group. Least challenging is the Hauptschule, which offers a five-year course to satisfy the requirement for compulsory schooling for the remaining fifty-five percent of an age group; most graduates become apprenticed in a trade or seek unskilled employment.

The schools we visited in April, 1974 were located in the university town of Göttingen, in north-central Germany. The Gymnasium and Hauptschule were coeducational, but the Realschule was in transition, with coeducational seventh grades but segregated ninth grades. (As we mentioned earlier, there were no regular eighth-grade classes in 1973-74). Our sample populations were taken from one of the two Gymnasiums in the city, from two Realschulen (one for boys and one for girls from the same attendance area), and two Hauptschulen (one near the city center, one in a recently built suburb occupied by workers in the expanding industrial section of the city).

The non-existence of regular eighth-grade classes was of serious concern to us. You can see from Table 1 that the average age of the German students was about a year greater than that of the Scandinavian, Italian, and American students, and about a half year greater than that of the British and Austrian students. To check the effect of age, we compared the scores of German ninth graders under fifteen years of age with those of their classmates over fifteen, and found virtually no difference between them. Since the ninth graders have only had three months of school more than eighth graders in other countries, we believe that our German principal group is adequate for the purposes of our study. Nevertheless, we sought to test a substantial pilot group of seventh graders to have some results for younger students. In the time available, we were able to administer both tasks to one coeducational seventh-grade Realschule class and the Control of Variables task to three additional Gymnasium classes.

The frequencies of the various categories of responses on the Proportional Reasoning and Control of Variables tasks are given in Figures 15 and 16. On the Proportional Reasoning task the Gymnasium students scored very well, with more than ninety percent giving Transitional or Ratio explanations. Boys and girls performed indistinguishably.

The Hauptschule boys and girls, who attended coeducational classes as well, also performed very similarly, though fewer than half used transitional or proportional reasoning. In the Realschule, however, we found very substantial and significant differences between the replies of boys and girls. The boys' performance was very close to that of Gymnasium students, while the girls' transitional and proportional answers were only sixty-six rather than ninety percent of the total, or at a level intermediate between that of the Gymnasium and Hauptschule students. The "rule-of-three" and rulers were much used here as in Austria.

On the Control of Variables task the results generally followed the same pattern. Gymnasium and Hauptschule students showed no significant sex differences, with about seventy percent of the Gymnasium boys and girls receiving a score of three points or more, while only four percent of the Hauptschule boys and girls gave explanations at this level. In other words, the Hauptschule students were almost at a total loss on the Control of Variables

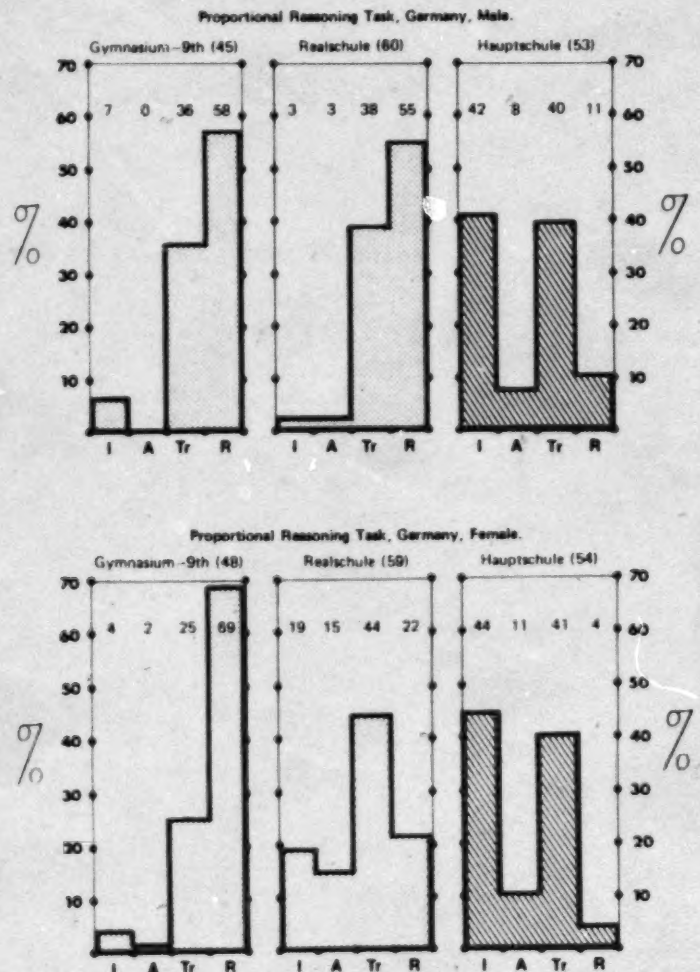


Figure 15. Proportional reasoning data, Germany.

task. The boys and girls in the Realschule again scored rather differently, with sixty-five percent of the boys scoring three or more (almost like the Gymnasium students), but less than twenty percent of the girls doing so. Here again the girls, from the same area in the city as the boys but attending a different school, were at a severe disadvantage.

Because of the age discrepancy of our German sample, we are reporting the score distribution of the seventh-grade students in the Gymnasium whom we were able to test on the Control of Variables task. These boys and girls, approximately one-and-a-half years younger than the ninth graders, scored very similarly to one another and about forty-five percent earned three points or more. In other words, they did not do quite as well as their older schoolmates, but many of them nevertheless showed a good understanding of the control of variables.

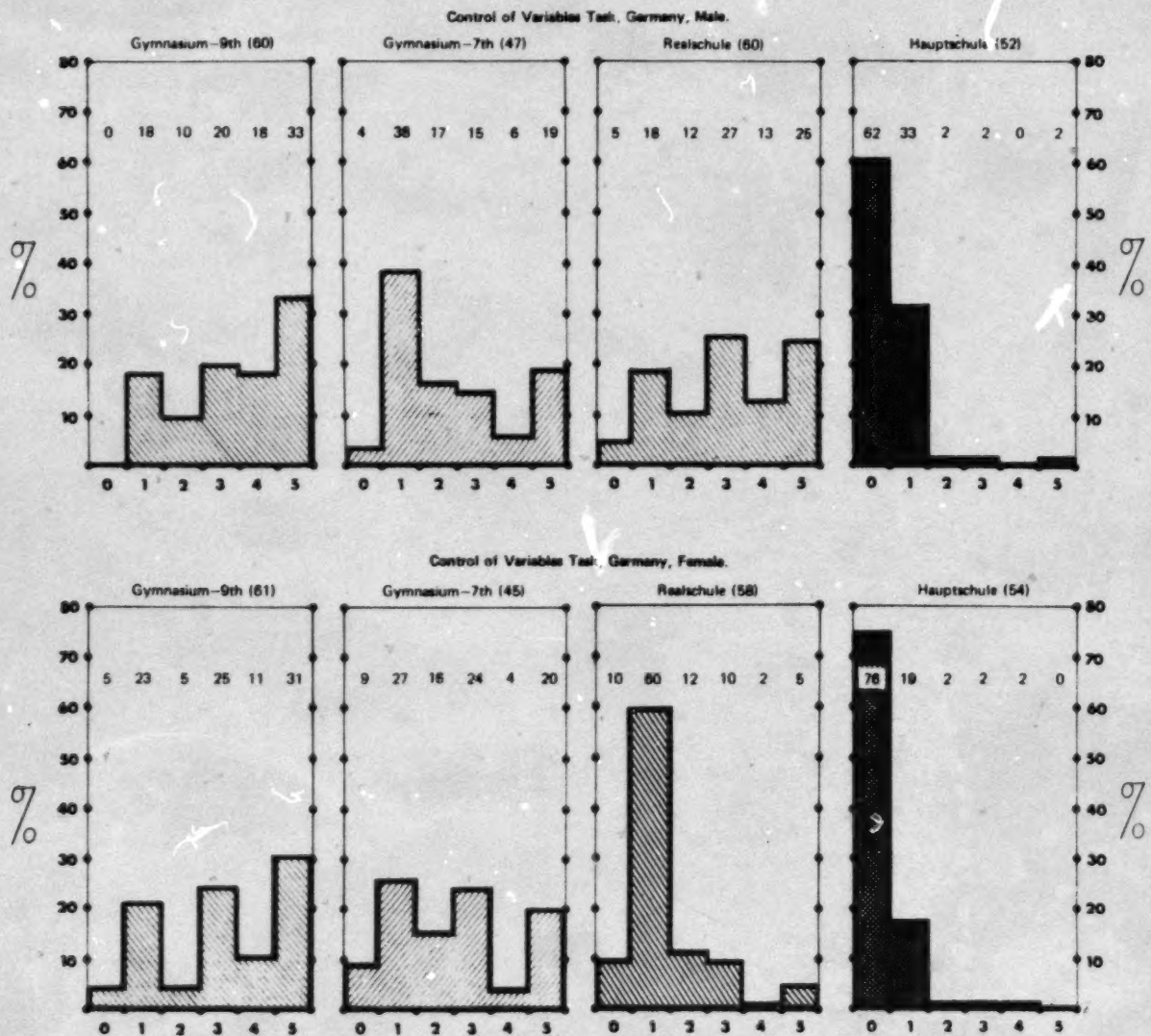


Figure 16. Control of Variables data, Germany.

The scattergram in Figure 17 presents the variation in student reasoning from class to class. The three classes for which we have only Control of Variables data are marked near the left axis; from the performance of similar students on the Proportional Reasoning task, we would expect them to have scored above seventy percent on that test. The trend here is for points to lie substantially below the diagonal. The very high performance of both seventh- and ninth-grade Gymnasium classes can be seen clearly, as can the very low performance of the Hauptschule classes on the Control of Variables task. The two classes of girls in the Realschule are represented by the two lower circles, while the two boys' classes are represented by the two high circles, very close to the three Gymnasium classes represented by squares. A seventh-grade coeducational Realschule class is also marked, and showed a somewhat better result than the ninth-grade girls' classes.

## Germany

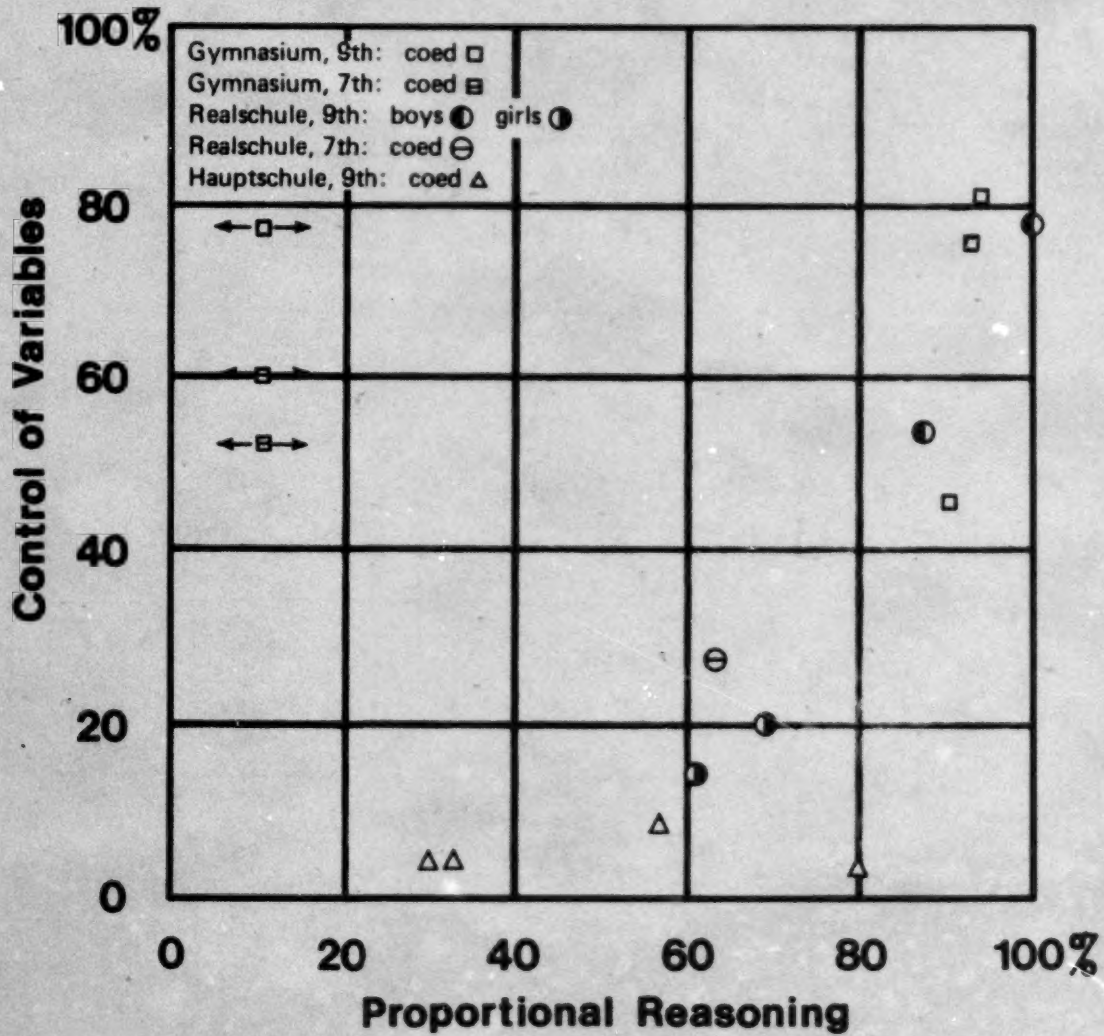


Figure 17. Scattergram of the percentage of German students in each class who showed transitional or formal reasoning on each task.

# RESULTS FOR GREAT BRITAIN

The British secondary school organization is quite complicated, depends somewhat on the community, and is gradually changing. We therefore proceeded somewhat arbitrarily, limiting our study to schools in London, and selecting sample populations from comprehensive schools, grammar schools (publicly administered), and direct grant schools (privately administered). The two latter types are highly selective, students in the top ten or fifteen percent of their age group as determined by an entrance examination at age eleven. The students we tested were completing their third year of secondary school, and averaged a little over fourteen years of age. The schools were located in various sections of the city, with a substantial minority of children of recent immigrants to London in the comprehensive schools. Very few students had difficulty with the language. There is academic tracking (called "streaming") of students in the grammar and comprehensive schools, but we do not know just how that affected our sample.

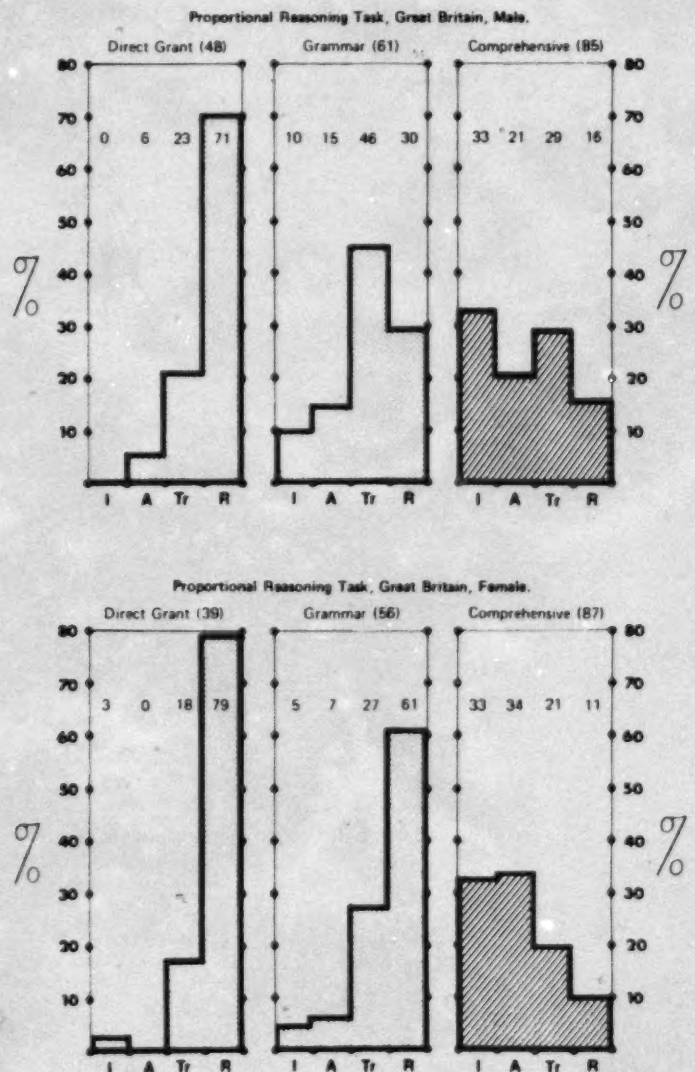


Figure 18. Proportional reasoning data, Great Britain.

The frequencies of the various categories of responses on the Proportional Reasoning and Control of Variables tasks are presented in Figures 18 and 19. The effect of student selection on both tasks is immediately evident, with the direct grant school students exhibiting almost completely transitional or formal thought on both tasks. About eighty percent of grammar school students gave Transitional or Ratio explanations on the Proportional Reasoning task, and somewhat over sixty percent scored three or more on Control of Variables. The comprehensive school students gave explanations that were substantially less advanced, with about forty percent Transitional or Ratio on proportional reasoning, and about thirty-five percent scoring three or more on the Control of Variables task. The differences between the three groups are statistically significant ( $p < 0.01$ ). Within each group, however, the only significant boy-girl difference was observed for proportional reasoning in the grammar school, with the girls providing more ratio explanations and the boys more Transitional ( $p < 0.01$ ). Since we sampled only one boys' and one girls' grammar school, and because of tracking within a school, this result must not be generalized.

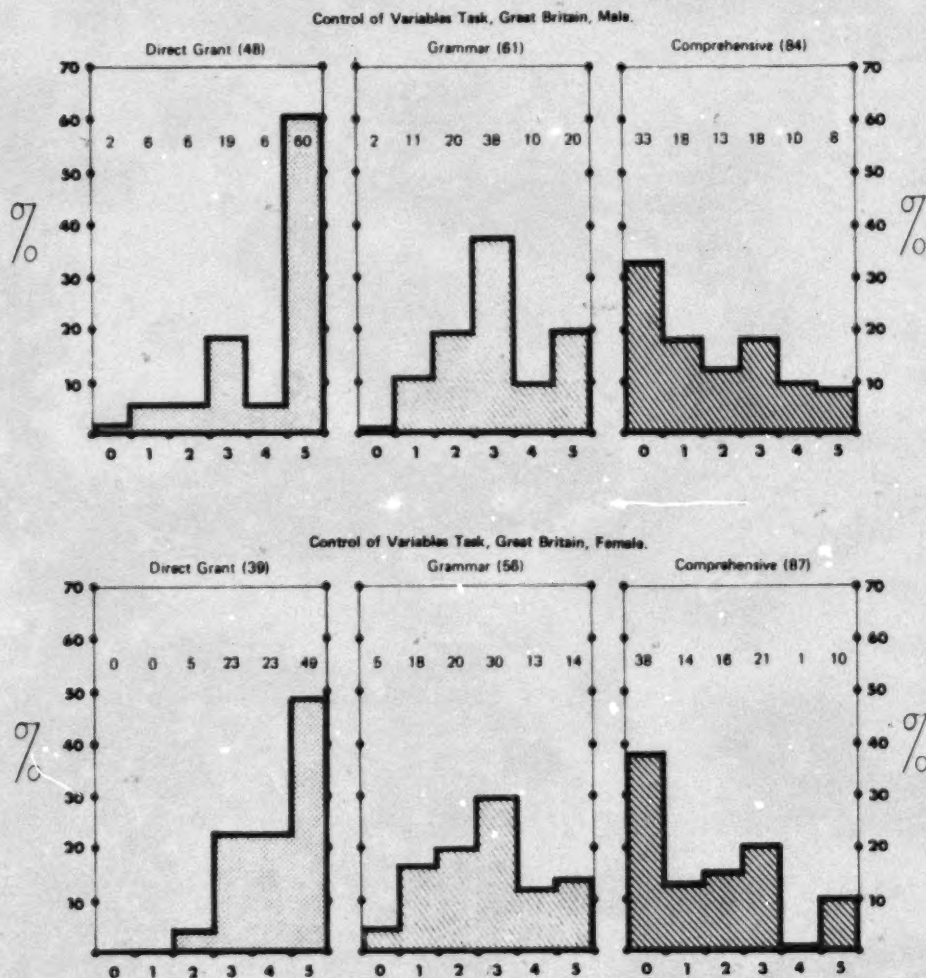


Figure 19. Control of variables data, Great Britain.

The enormous range of reasoning from class to class is illustrated in the scattergram in Figure 20 where the points lie near the diagonal but are spread from one end to the other. Thus, the two tasks were of comparable difficulty. The three school types are neatly separated, except for one outstanding comprehensive school class that ranks with the direct grant schools in this respect. We suspect that tracking is responsible, since the other classes in the same school did not give such a performance.

## Great Britain

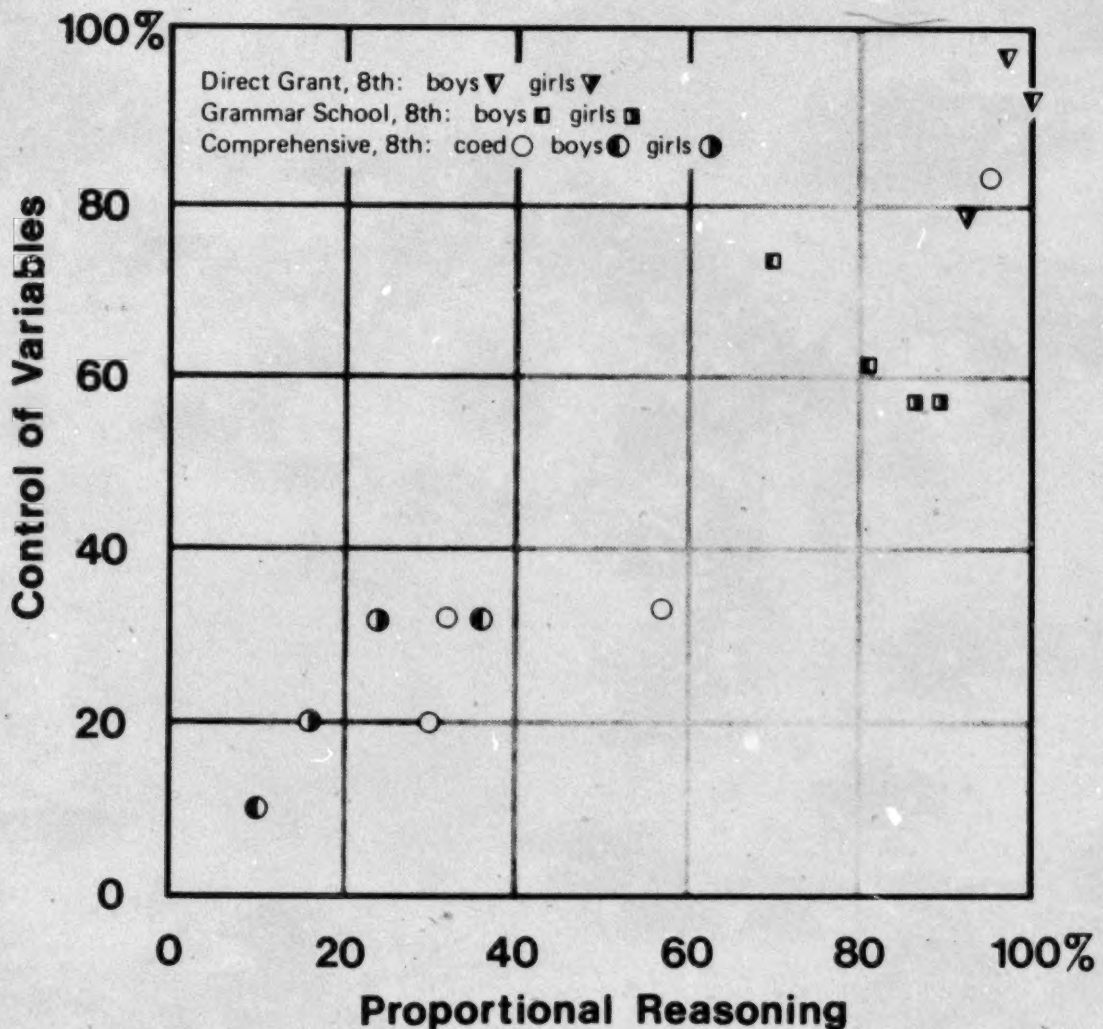


Figure 20. Scattergram of the percentage of British students in a class who showed transitional or formal reasoning on each task.

## COMPARISONS OF POPULATIONS

Because of the complexity of the data, it is difficult to get a comparative overview of the thirty-eight populations (nineteen each, male and female) for which we have reported results in the preceding sections. To help in this effort, we have looked for patterns in the frequency distributions in Figures 2 to 20 and have grouped the populations into six sets with respect to each of the two tasks. The patterns identifying each set are listed in Table 7. We have numbered the patterns in order of decreasing evidence of formal thought (or increasing evidence of intuitive thought). Thus, Pattern I in each case includes the highest percentages of formal thought, Pattern II somewhat less, and so on. You will notice that certain patterns for the Proportional Reasoning task have been given the same number because they represent closely equal average levels of reasoning and differ primarily in the frequency of occurrence of Additive responses. Table 7 also indicates how we have distinguished among the various patterns by using differing designs in the graphs of the earlier figures.

We should like to point out a few features of the patterns we have selected. You will note that responses in the Ratio category to the Proportional Reasoning task are the mode in Pattern I, but are important also in Patterns IIA and IIB. Intuitive responses are prominent in all patterns except I, and make up more than half of the responses in Pattern IV. Additive explanations are substantial only in Patterns IIB and IIIB. The patterns on the Control of Variables task show a more gradual change from formal thought in I to intuitive thought in VI, as the modal score decreases from 5 to 0 points.

The assignment of populations to the patterns is given in Table 8. We have entered the letter M in a column to represent the performance of the male population listed at the left, and have entered F for the female population. The populations themselves are grouped according to whether the educational systems are comprehensive or selective, and then by socio-economic criteria or school type.

The overview afforded by Table 8 reveals several interesting results. Most striking is the uniformly high degree of proportional reasoning manifested by the top academic groups in the selective school systems. This achievement extends to the boys but not to the girls in the academic middle groups. The populations in comprehensive school systems, as might be expected, tend to fall into the intermediate patterns with respect to proportional reasoning. Only the Italian upper middle class prep school boys--actually a selected group--excel in formal thought on this task, while the urban low income groups in the United States appear to lack formal thought completely. The United States populations, whose best achievement is Pattern IIB, do not compare favorably with the approximately equivalent European groups, especially European boys. Many girls' preference for Additive responses, already mentioned in connection with the results for several countries, can be seen again here, in that the distributions of nine girls' populations fell into the high Additive Patterns IIB and IIIB, while only five boys' populations showed these distributions.

On the Control of Variables task, the differences in selectivity in the participating countries become more evident. The most selective group, the British direct grant population, is clearly far superior to any other population. The relatively less selective Austrian Gymnasium is at a disadvantage

compared to the German Gymnasium and the British grammar schools, with the eighth-grade students achieving at the same level as German seventh graders who were about a year younger. Still, the Austrian third quartile (Hauptschule A) equaled the British comprehensive school students who make up eighty percent of the population; the Austrian fourth quartile (Hauptschule B), which could virtually not cope with the Control of Variables task at all, shared this difficulty with the German Hauptschule (third and fourth quartiles).

The students in comprehensive school systems succeeded rather better on this task than they had on proportional reasoning. The upper middle class children in the United States achieved Pattern II, equal to or better than the selected top track students with the exception of the direct grant group. The Italian and Scandinavian students equalled the top or middle track students in Germany, Great Britain, and Austria, rather than equalling the lowest track students as they had on proportional reasoning. Even the boys in the urban low income group in the United States achieved in Pattern V on Control of Variables, whereas they had been able to do virtually nothing with the Proportional Reasoning task.

Table 7 Frequency Distribution Patterns

| Patterns on Proportional Reasoning Task                                             |                                                                                    |
|-------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
|    | I Predominantly Transitional and Ratio                                             |
|    | IIA Predominantly Intuitive, Transitional, and Ratio, very few Additive*           |
|  | IIB Even distribution in all categories*                                           |
|  | IIIA Predominantly Intuitive and Transitional, few Additive and Ratio <sup>+</sup> |
|  | IIIB Predominantly Intuitive, Additive, and Transitional, few Ratio <sup>+</sup>   |
|  | IV Predominantly Intuitive                                                         |
| Patterns on Control of Variables Task                                               |                                                                                    |
|  | I Almost exclusively 3-5 points                                                    |
|  | II Fairly even distribution with an excess of scores 3-5                           |
|  | III Fairly even distribution with an excess of scores 0-2                          |
|  | IV Modal score 0 or 1, with some scores of 2-5 points                              |
|  | V Predominantly 0 and 1 points                                                     |
|  | VI Predominantly 0 points                                                          |

\*Categories IIA and IIB are closely equal in their "average" level of reasoning, but the distributions are strikingly different.

<sup>+</sup>Categories IIIA and IIIB are closely equal in their "average" level of reasoning, but the distributions are strikingly different.

Table 8 Comparison of Populations on the Two Tasks

| Population                            | Frequency Distribution Pattern |     |     |      |      |     |                      |     |     |     |     |     |
|---------------------------------------|--------------------------------|-----|-----|------|------|-----|----------------------|-----|-----|-----|-----|-----|
|                                       | Proportional Reasoning         |     |     |      |      |     | Control of Variables |     |     |     |     |     |
|                                       | I                              | IIA | IIB | IIIA | IIIB | IV  | I                    | II  | III | IV  | V   | VI  |
| <b>Comprehensive School Systems</b>   |                                |     |     |      |      |     |                      |     |     |     |     |     |
| <u>Upper Middle Class</u>             |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Italy                                 | M                              |     | F   |      |      |     |                      |     | M   | F   |     |     |
| United States                         |                                |     | M,F |      |      |     | M,F                  |     |     |     |     |     |
| <u>Middle Class</u>                   |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Sweden                                |                                | M,F |     |      |      |     | F                    | M   |     |     |     |     |
| Italy                                 |                                | M   |     |      | F    |     |                      | M,F |     |     |     |     |
| United States                         |                                |     |     |      | M,F  |     |                      | M,F |     |     |     |     |
| Denmark                               |                                | M   |     | F    |      |     |                      |     |     | M,F |     |     |
| <u>Working Class &amp; Low Income</u> |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Sweden                                |                                |     |     |      | M,F  |     |                      |     | M   | F   |     |     |
| Italy                                 |                                | M   |     | F    |      |     |                      |     |     |     | M,F |     |
| United States                         |                                |     |     |      |      | M,F |                      |     |     |     | M   | F   |
| <b>Selective School Systems</b>       |                                |     |     |      |      |     |                      |     |     |     |     |     |
| <u>Top Groups</u>                     |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Direct Grant (Great Britain)          | M,F                            |     |     |      |      |     | M,F                  |     |     |     |     |     |
| Grammar (Great Britain)               | M,F                            |     |     |      |      |     |                      | M,F |     |     |     |     |
| Gymnasium (9th, Germany)              | M,F                            |     |     |      |      |     |                      | M,F |     |     |     |     |
| Gymnasium (7th, Germany)              |                                |     |     |      |      |     |                      |     | M,F |     |     |     |
| Gymnasium (Austria)                   | M,F                            |     |     |      |      |     |                      |     | M,F |     |     |     |
| <u>Middle Groups</u>                  |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Realschule (Germany)                  | M                              |     | F   |      |      |     |                      | M   |     | F   |     |     |
| Hauptschule A (Austria)               | M                              |     | F   |      |      |     |                      |     |     | M   |     | F   |
| Comprehensive (Great Britain)         |                                |     |     |      | M,F  |     |                      |     |     | M,F |     |     |
| <u>Low Groups</u>                     |                                |     |     |      |      |     |                      |     |     |     |     |     |
| Hauptschule (Germany)                 |                                |     |     | M,F  |      |     |                      |     |     |     |     | M,F |
| Hauptschule B (Austria)               |                                |     | M   |      | F    |     |                      |     |     |     |     | M,F |

A comparison of the students' relative performance on the two tasks is made most easily with the help of the seven scattergrams in Figures 4, 5, 8, 11, 14, 17, and 20. One can see that the United States is the only country for which the points fall generally above the diagonal, indicating that the Proportional Reasoning task was more difficult than the Control of Variables task as evaluated by our scoring procedure. For England and Sweden the points lie relatively close to the diagonal and give evidence that the tasks were of comparable difficulty both for higher and for lower achieving groups. For Denmark and Italy the points are distributed broadly below the diagonal, which means that the Control of Variables task was somewhat more difficult than the Proportional Reasoning task, but that the difference was not very great. In Germany and Austria, finally, the points tend to lie near the bottom and right edges of the diagrams; the implication is that controlling variables was very hard to understand for all but some of the most able students.

This last observation raises questions about the adequacy of the translation of the Control of Variables task into German. We have already explained that the language on the answer pages gave many children difficulty, so that they asked for supplementary oral clarification. The best way to resolve this concern would be to use slightly differing forms of phrasing the task and then to determine how greatly these differences affect the students' performance. As a matter of fact, trials of this kind did precede our designing of the questions in their original English forms. Unfortunately, our travel schedule did not permit us to carry out similar pilot studies in each of the countries visited.

In the absence of such an investigation, we have looked for clues within our present data. Consider first the German boys and girls attending the Realschule. Whatever language problems there might be, both groups were certainly affected in the same way. The substantial score difference observed with respect to both tasks (see Figures 15-17) can only be explained in terms of student selection by the schools, the teaching program, and sex-related cultural biases. Next, we ask whether any German-speaking groups performed without apparent language handicap on the Control of Variables task. The answer is yes the ninth-grade German Gymnasium students and the ninth-grade Realschule boys scored according to Pattern II, as did the approximately comparable upper middle class students in the United States and the British grammar school students. Furthermore, the Austrian Gymnasium students and the German seventh graders scored in Pattern III like the middle class students in the United States and in Italy, and somewhat better than the comprehensive school students in London. In other words, these German-speaking subjects performed quite adequately, though not up to their exceptional achievement on the Proportional Reasoning task. We therefore conclude that the translation did not introduce any substantial bias into the data, and that the very great difficulty experienced by the Hauptschule students in Austria and Germany is due to student selection, emphases of the teaching program, and possibly cultural biases.

## DISCUSSION AND CONCLUSIONS

For the sake of clarity, we shall divide this section into four parts dealing, respectively, with proportional reasoning, control of variables, implications of both tasks insofar as formal thought is concerned, and the comments and opinions expressed by the subjects.

Proportional Reasoning. The great diversity of methods used by our subjects to find the height of Mr. Tall is similar to what we had observed previously in schools of the San Francisco Bay Area.<sup>7,8</sup> The second question concerning the width of the car had not been used before. It is interesting to note, therefore, that about two-thirds of the students used the same method on both questions (Table 2). Most of the remainder combined methods in one of two ways: (1) the Additive procedure for Mr. Tall, followed by an Intuitive method--often scaling, as in Examples I9 and I10--for the car; (2) Ratio for Mr. Tall, followed by a Transitional approach for the car. An additional five percent of subjects used Ratio (R) for Mr. Tall followed by Intuitive (I) for the car. We believe that most students in this small though not negligible group had learned an algorithm for proportions that they could apply to the direct first question but could not invert to answer the second question. Lack of time, loss of interest, and possibly copying of the first answer may also have contributed to the R-I response pairs.

As we have pointed out before, the car question was more difficult than the question concerning Mr. Tall and had the intended effect of further challenging the students to reveal how thoroughly they could apply proportional reasoning. Our procedure for constructing the composite score used in most of the figures and tables in this article, therefore, has made this assessment of proportional reasoning more stringent than it was in our earlier studies.

We were repeatedly struck by the similarity of individual responses, which occurred again and again and could be recognized clearly regardless of the language used. Even though the frequency distributions varied rather widely, we believe that all the population groups we sampled share a basically similar development of their patterns of reasoning. As a matter of fact, this conclusion is not really surprising, since all the countries we visited are culturally and technologically similar.

Most of the explanations in categories A and R were simple and straightforward, and could be identified easily. A few of the Additive replies explicitly mentioned that paper clips and buttons were about the same size to justify the procedure (see Example A2), but of course did not take into account the size difference indicated by the two measurements of Mr. Short. Thus they did not coordinate all the known data.

The explanations in categories I and Tr showed a great deal of variety, and we chose the illustrative examples to communicate this variety. Because of the contradictions they contain, the misconceptions in many replies could be interpreted only with difficulty (see Example Tr1; did the student add six to one-and-one-half instead of multiplying, or did he mean to add two times one-and-one-half for the two additional buttons?). In Category I, for instance, there are some guesses (Examples I1 and I7), operations with numbers regardless of their significance (Example I4), introduction of extraneous data (Examples I2 and I8), and unexplained scale relationships (Examples I3, I5, I9, I10, I11). Finally, there were a few subjects who believed that the problems could not be solved without their seeing Mr. Tall, the car, and/or the buttons (Example I6).

One recurring pair of responses classified A-I concluded that the car was seven buttons wide because that was half the paper clip measurement of fourteen, very possibly using division by two as the inverse operation to the addition of two. There were even some explanations in Categories I-I that stated "two times as big in paper clips" for Mr. Tall, but then added  $2 + 6 = 8$  rather than multiplying  $2 \times 6 = 12$ .

Included in the Transitional category is a substantial number of responses that determined the conversion from buttons to paper clips by using a ruler to measure Mr. Short and the paper clips, then expressing all dimensions in inches or centimeters. This procedure occurred primarily in Austria and Germany, where each pupil carried at all times a small supply kit containing pens, pencils, rulers, and erasers. Even though using a ruler was easily visible to neighboring students, this approach was not copied very often. Another group of responses we included in the Transitional category started out with  $7\frac{1}{2}$  for the height of Mr. Short in paper clips (probably a counting error), and concluded that Mr. Tall was about twelve paper clips high because each button was about two paper clips. Since we had previously found that use of a two-to-one ratio is not indicative of proportional reasoning at the level of formal thought,<sup>3</sup> we felt that it would not be appropriate to include these responses in Category R.

The rather large percentage of students replying in the Additive category was originally one of the surprising results of the Proportional Reasoning task.<sup>7,8</sup> Wollman suggested that this effect might be induced partially by the 2:3 ratio of heights, which acts to discourage comparison by dividing and draws particular attention to the excess of one dimension over the other.<sup>21</sup> In Appendix B we have described a small study using the car question before the Mr. Tall question, and showed that the latter question did appear to influence about twenty percent of the subjects. It was therefore interesting to observe that the composite Additive responses occurred at a frequency of more than twenty percent in fourteen of the thirty-six populations we investigated, in five of the seven countries. Only Germany and Denmark lacked such populations entirely, with the percentages there ranging from zero (German Gymnasium, boys) to fifteen (German Realschule, girls).

All six United States populations had large frequencies of Additive responses, ranging from seventeen to forty-one percent. The Austrian groups divided rather curiously: three populations had between three and five percent Additive responses, while the other three gave between twenty-six and thirty percent. In Italy the upper middle class and middle class girls gave more than twenty percent Additive replies, while the corresponding boys' groups and the working class students contributed low percentages. In Sweden and Great Britain, two groups of lower socioeconomic or ability levels gave more than twenty percent Additive answers.

To what extent are the Additive explanations developmentally intermediate between Intuitive and more advanced procedures? Clearly the frequency distributions in Pattern I (see Table 7) reflect such advanced proportional reasoning that few Additive answers would be expected. Conversely, frequency distributions in Pattern IV indicate so little development that one would also expect relatively few Additive responses. Patterns II and III are therefore the important ones to consider.

The present data strongly suggest that additive reasoning does not lie on an invariant developmental sequence but is strongly influenced by instruction and represents an effort by students to deal with a task in an ad hoc rather than systematic way. Thus, the German Hauptschule students, the Danish students, the Swedish middle class students, and the Italian middle and working class boys all show frequency distribution Patterns IIA or IIIA (see Table 8), which have substantial Intuitive and Transitional contributions but lack Additive ones. In fact, all of these groups have a larger percentage of individuals responding Additive-Intuitive to the two questions on the Proportional Reasoning task than respond Additive-Additive.

At the same time, the Swedish working class students, the American upper middle and middle class students, the Italian upper middle and middle class girls, the Austrian Hauptschule-B students and Hauptschule-A girls, and the British comprehensive school students all had frequency distribution patterns IIB or IIIB, which included more than fifteen percent of Additive responses and a comparable percentage of Additive-Intuitive answer pairs. Most of these students attend different schools and therefore may experience different courses of study than the student groups in the same countries that gave Additive responses. It appears that the educational programs in the United States, Austria, and Great Britain tend to encourage Additive procedures while those in Germany and Denmark tend to discourage it on the part of the students who have not yet mastered proportional reasoning. In Italy and Sweden the results are mixed, depending on sex or socioeconomic level.

Control of Variables. Our scoring procedure for the three questions on this task was simple to apply but partially obscured the qualitative nature of the answers given by an individual. The application of the concept of a controlled experiment occurred most frequently on Question 3, somewhat less frequently on Question 2(b), and least frequently on 2(a).

Many responses to Question 2(a) receiving zero score dealt with the positions and actions of the two spheres independently (Examples 1 and 4), while others used the light sphere's height to compensate for the heavy sphere's weight (Examples 2 and 3). Zero score explanations to Question 2(b) tended to relate to the heavy sphere only (Examples 6 and 7) or communicated the need to maximize the difference in target distances (Example 9). All of these answers--many of them quite lengthy and carefully stated--suggest that the respondent was not conceptualizing the experimental comparisons that were described in the questions.

In other words, rather than considering the effect of the weight on target displacement, many subjects focussed on the effect of the sphere on the target position. The former point of view addresses a relationship between two abstract variables (weight and displacement) which requires a formal operation; the latter concerns a relationship between two tangible variables (sphere and position) which requires only a concrete operation. In the concrete approach there is no need for two spheres, hence the two trials in Question 2(a) are seen as unrelated.

Scores of one point on Questions 2(a) and (b) were awarded for explanations that identified the equality of starting positions as important but not as necessary. Here also there may have been a semantic problem, in that the subjects were asked to explain their ideas but were not urged, as would have happened during an interview, to give the strongest possible justification. These intermediate replies, therefore, reflect the subject's spontaneous view of how well-grounded an explanation was required.

The maximum score of two points was given for explanations that indicated the necessity of equal starting points for the two spheres. As indicated in Table 6, only about twenty percent of the students earned a total score of 4 or 5 by doing this consistently on all questions. And most of these merely asserted that equality was necessary (Examples 1 to 5), but did not refer to the ambiguities that would arise if the starting positions were unequal (Example 6). Some of them did, however, refer to the more general principle that only one variable at a time could be changed (see Examples 1, 2, 4, 5). Again, an interview would have allowed for probing questions after the first answer, while the group administration of the task was limited to the three standard questions printed on the answer page.

Question 3, though identified as an inadequate comparative experiment by about two-thirds of all students, was not well understood by others who were given zero points. Some explanations, for instance, seemed to concern the reason why metal A hit the target farther (Example 2). Others brought in extraneous information (Examples 4, 5, and 6), much like some of the Category I answers to the Proportional Reasoning task, while still others referred to the concrete happening described in the question (Example 1).

Of the explanations that earned one point, only a minority referred to both starting position and target weight as uncontrolled variables (Examples 1 and 4). The others either mentioned only the position (Example 2) or the target weight (Example 3). In addition, many supplied comments regarding the mechanism by which the position or target weight would have its effect (Example 2).

It would seem that successful design and analysis of experiments depend on an individual's ability to recognize multiple causes. Question 1 of the Control of Variables task asked the students to enumerate the variables that might affect how far the target goes. Virtually all students named some variables, and many participated in the brief discussion that preceded the remainder of the task. The quality and number of the written items in answer to Question 1 varied greatly from student to student; very few appeared to add to their list during the discussion.

Some of the variables frequently mentioned could be classified as experimenter manipulated, such as the rolling sphere's weight and starting height, the target weight, and the angle of the ramp. Other variables, such as the sphere's striking speed and force of impact, could not be directly manipulated by the experimenter. Some students listed objects (the sphere, the track) and others made descriptive statements (the heavy sphere will hit it hard). The items given bore little relationship to the answers to Questions 2 and 3. We therefore conclude that designing a controlled experiment requires substantial logical insight beyond an awareness of variables, and that many eighth graders lack this insight.

Formal Thought. Our findings support and add further detail to the findings of others that the development of formal thought among adolescents is at best a gradual and incomplete process.<sup>8,12,22-26</sup> As we have pointed out, the nature of our tasks, with their lack of feedback or encouragement for the subjects, provides a rather stringent assessment. Piaget's clinical interview characteristically proceeds as a dialogue in which the subject gradually gains information from the investigator or by using a piece of apparatus, and one can observe how he responds to the new data. In such a situation, many subjects may ultimately show reasoning patterns more advanced than we were able to collect. Yet, in an earlier study of proportional reasoning that compared interviews and group paper-and-pencil tasks, we found that the differences were moderate.<sup>8</sup> Hence we believe that the main features of our findings have general significance. Furthermore, our tasks are rather similar to what is expected of secondary school students in science and mathematics classes.<sup>27,28</sup> Literary and historical analysis also, though qualitative rather than quantitative, depend on students' abilities to evaluate evidence, identify causes, and marshal a conclusive argument, and therefore require reasoning patterns like those evaluated by our tasks.

Formal thought in our study manifes itself as fluent use of ratio on the Proportional Reasoning task (composite score R) and a score of five points on the Control of Variables task. We actually found 251 such students, or about seven percent of our sample. At the other extreme, there were 422 students, or about fourteen percent, who used intuitive reasoning on the Proportional Reasoning task and scored zero points on control of variables. Both groups included about equal numbers of boys and girls.

To show how reasoning on the two tasks was correlated, we present the contingency table in Table 9. To give equal weight to all countries in spite of differences in sample sizes, we have again used the averages of the percentages for the fourteen populations (seven countries, two sexes) as we did in Tables 2, 4, 5, and 6. A close examination of Table 9 shows that most of the subjects who scored three, four, or five points on the Control of Variables task responded in Categories Tr or R on proportional reasoning, an average of twenty-eight percent of a population. This is the level of performance that we presented in the scattergrams for the individual countries and identified as being at least transitional to formal thought.

It is also clear from the table that very few of the students who scored zero points on control of variables answered in the Ratio category on proportional reasoning. Subjects with one or two points, however, were approximately equally likely to respond in any one of the four categories I, A, Tr, or R. The failure of the two scores to correlate more closely reflects once again the fact that students in several countries found the Control of Variables task more difficult and scored only one or two points even though they gave evidence of proportional reasoning.

The features we have just described are summarized in Table 10 where we have grouped categories I with A, Tr with R, scores 1 with 2, and scores 3 with 4 and 5. Each of the resulting six entries for the average percentage is followed, in parentheses, by the expected figure calculated on the assumption that the two tasks are completely uncorrelated. The lowest and the highest groups on control of variables are substantially correlated with the corresponding performance on proportional reasoning, but the middle group (scores of 1 or 2) is not. It is this lack of correlation for the middle group and the great differences in correspondence of the two scores from country to country that casts doubt on the existence of a single operational structure underlying all manifestations of formal thought. Rather, it would seem that proportional reasoning can be substantially advanced through the teaching programs used in some of the countries we visited, while control of variables can be enhanced through others.<sup>29-32</sup>

Our findings also have implications for teachers. According to Tables 9 and 10, about thirty-seven percent of eighth-grade students lack formal mental structures in both proportional reasoning and control of variables, and thirteen percent appear to be non-operational in these areas (Category I and score 0). Another thirty-six percent have formal structures in only one of the areas, either proportional reasoning (thirty percent) or control of variables (six percent). It would seem, therefore, that teaching goals and teaching

Table 9 Contingency Table for Composite Scores on the Proportional Reasoning and Control of Variables Tasks (percent)

|                              |       | Control of Variables Score |    |    |    |   |    | Total |
|------------------------------|-------|----------------------------|----|----|----|---|----|-------|
|                              |       | 0                          | 1  | 2  | 3  | 4 | 5  |       |
| Proportional Reasoning Score | I     | 13                         | 7  | 3  | 2  | 0 | 1  | 27    |
|                              | A     | 6                          | 4  | 2  | 2  | 0 | 1  | 15    |
|                              | Tr    | 6                          | 8  | 5  | 5  | 2 | 4  | 30    |
|                              | R     | 2                          | 6  | 4  | 5  | 3 | 8  | 27    |
|                              | Total | 26                         | 26 | 14 | 14 | 5 | 14 | 100   |

Table 10 Condensed Contingency Table for Scores on the Two Tasks (percent)

|             |       | Control of Variables Score |         |         | Total     |
|-------------|-------|----------------------------|---------|---------|-----------|
|             |       | 0                          | 1,2     | 3,4,5   |           |
| Prop. Reas. | I,A   | 19 (11)*                   | 18 (17) | 7 (15)  | 43 (43)   |
|             | Tr,R  | 8 (15)                     | 22 (23) | 27 (19) | 57 (57)   |
|             | Total | 26 (26)                    | 40 (40) | 34 (34) | 100 (100) |

\*Figures in Parentheses are percentages expected in the absence of correlation.

practice might well recognize these differences and address themselves more consciously to the task of building formal operational structures.<sup>33-35</sup> During our school visits we became aware of some efforts by individual teachers in this direction, but did not observe any major, concerted activity. The syllabi and most of the textbooks in general use seemed poorly suited to this task.

A very interesting question raised by our study concerns the effects of selective versus comprehensive school systems as regards formal thought. The scattergrams in Figures 4, 5, 8, 11, 14, 17, and 20 summarized the approximate level of formal thought encountered by a science or mathematics teacher in various class groups. We have already pointed out the extent to which selectivity or socioeconomic level influenced these graphs. In the absence of large-scale longitudinal studies we cannot distinguish the impact of specific teaching programs from the effects of original selection. We are not convinced of the superiority of either approach and believe that improvements could be made within both. In our opinion, the most urgent problems derive from the difficulties of the low-achieving groups and their needs for learning activities strongly rooted in concrete operations.

Student Opinions. At the conclusion of the two tasks, the subjects were invited to use the backs of their answer pages for describing their opinions of the research project and our visit to their class. About two-thirds of them did so. The vast majority expressed positive attitudes, but a handful out of about three thousand students were dissatisfied because they disliked tests, had not been consulted in advance, or felt they were being exploited. Students liked the visit because it allowed them to avoid their regular assignments, challenged them in unusual ways, enabled them to participate in an international activity, and/or acquainted them with new scientific ideas. Many wished us a good trip home (they had been impressed by our coming from California to work with them) and hoped we would come again.

Some students indicated that their teachers needed to be more tolerant of a diversity of approaches to a problem; many German Gymnasium students, however, expressed surprise at our project because they considered it obvious that different persons would use different methods for solving the tasks. A few subjects expressed curiosity about our findings and asked to see our report, and some even pointed out the difficulty we faced of adequately sampling students in their country. Quite a few students confessed that they did not really understand the purpose of our project. Here are a few comments quoted from English language answer pages:

I think that it was an unusual experience, but it helped me to think before I try something. I think it should be done more often to help students understand things like math, and so forth, better.

I like science and the teacher, but sometimes science is difficult, so keep on coming.

This visit was very considerate of you and very nice. I am sorry you all could not see my best work. My head isn't together today. One day I wish to be a 3rd grade teacher and I am going to work on up until I make it. I really enjoyed your visit. Goodbye for now.

Since I understand why you came here I cooperated. It is a survey to help teachers understand their students better.

This experiment of having us write our ideas is good because the teachers need to know how their students' ideas vary and that not everyone thinks the same.

I don't know why you made these tests, but I think they were good. Because they got me thinking on how you can figure out questions and problems with your own head and not the teacher. I liked the sphere problem. I would have liked to use the metal right in front of me, but on the paper was alright.

To Dr. Karplus: I, . . . myself have really enjoyed you here today. And I have really learned how to say and write what I and only I think, with no one else to try to tell me. So I hope you come back again.

I would like to see the results of the experiment.

I really appreciate your coming. I might be at college some day.

It was nice to have Mr. Karplus here. The papers he gave us was hard and not understanding for me, but I think I will understand sometime when I go to college.

I didn't mind your testing. It was different from anything else I've done.

I think it was good because the teachers will have some interesting problems to teach and they'll know that there isn't only one answer to a problem.

My opinion of this test is that I don't mind it. I found it reasonably easy. Anyway, it got us off biology.

Even though the experiments were quite simple, it was hard to write down your thoughts on paper. In general it was quite hard.

It was a good change from our usual science lesson. It was interesting and made you think.

Very good idea.

It makes an interesting change to have tests like these once in a while, tests which are interesting with a hint of problems. It also makes a change for visitors to take a lesson. This gives the lesson a mysterious hint about it (you never know what's going to come next). Very interesting!

I thought the first paper was quite hard because the man in the picture was standing with his legs apart which made him much shorter.

I hope your research helps teachers in teaching people.

# References

1. B. Inhelder and J. Piaget, The Growth of Logical Thinking from Childhood to Adolescence (New York: Basic Books, 1958).
2. E. A. Lunzer and P. D. Pumfrey, "Understanding Proportionality," Mathematics Teaching, 34 (1966): 7-12.
3. K. Lovell and I. B. Butterworth, "Abilities Underlying the Understanding of Proportionality," Mathematics Teaching, 37 (1966): 5-9.
4. K. Lovell, "Proportionality and Probability," Piagetian Cognitive Development Research and Education, edited by M. F. Roszkopf, L. P. Steffe, and S. Taback, Reston, VA: National Council of Teachers of Mathematics, 1971.
5. K. Lovell, "A Follow-Up Study of Inhelder and Piaget's 'The Growth of Logical Thinking'," Brit. J. Psych. 52, 2 (1961): 143-153.
6. L. Kohlberg and C. Gilligan, "The Adolescent as a Philosopher: The Discovery of the Self in a Postconventional World," Daedalus, 100 (Fall 1971): 1051-1084.
7. R. Karplus, E. Karplus, and W. Wollman, "Intellectual Development Beyond Elementary School IV: Ratio, The Influence of Cognitive Style," School Science and Mathematics, 74 (October 1974): 476-482.
8. W. Wollman and R. Karplus, "Intellectual Development Beyond Elementary School V: Using Ratio in Differing Tasks," School Science and Mathematics, 74 (November 1974): 593-611.
9. W. Wollman, "Intellectual Development Beyond Elementary School VI: Controlling Variables: A Survey," School Science and Mathematics, submitted.
10. Antonio Suarez, "Die Entwicklung der Denkopoperationen beim Verständnis Funktionaler Zusammenhänge," unpublished doctoral dissertation, Eidgenössische Technische Hochschule, Zurich, 1974.
11. B. Andersson, "Some Simple Experiments Illustrating School Children's Entry into the Stage of Formal Thought," Kognitiv Impuls, Gothenburg: School of Education, 1974.
12. A. E. Lawson and J. W. Renner, "A Quantitative Analysis of Responses to Piagetian Tasks and Its Implications for Education," Science Education, 58, 4 (1974): 545.
13. D. Kuhn, J. Langer, L. Kohlberg, and N. Hahn, "The Development of Formal Operations in Logical and Moral Judgment," The Journal of Genetic Psychology monograph, in press.
14. R. Case, "Structures and Strictures: Some Functional Limitations on the Course of Cognitive Growth," Cognitive Psychology, 6 (1974), 544-573.
15. L. C. Comber and J. P. Keeves, Science Education in Nineteen Countries (Stockholm: Almqvist & Wiksell, 1973).

16. See reference 15, Appendix IX; the items are on Test 1A, #8, 9, 14; on Test 1B, #10, 18; on Test 10A, #3, 4, 10, 11; on Test 11B, #4, 6, 12.
17. M. M. Downie and R. W. Heath, Basic Statistical Methods, 3rd Ed. (New York: Harper and Row, 1970): 123-124.
18. Our scoring procedure differs somewhat from that used in Reference 9.
19. See reference 17, 124-125.
20. Marina Pascucci (Formisano), "Sviluppo Cognitivo e Pensiero Formale, Primi Risultati di una Ricerca Sperimentale Interculturale," Tesi di laurea in Pedagogia, Università degli Studi di Roma, 1973/74, contains a more extended description of the effects of sex, mixed vs. segregated classes, and socio-cultural variables.
21. Warren Wollman, private communication.
22. D. Elkind, "Quantity Conceptions in Junior and Senior High School Students," Child Development, 32 (1961): 551.
23. K. Lovell, "Some Problems Associated with Formal Thought and Its Assessment," Measurement and Piaget, edited by D. R. Green, M. P. Ford, and G. B. Flamer (New York: McGraw-Hill Book Co., 1971).
24. J. Piaget, "Intellectual Evolution from Adolescence to Adulthood," Human Development, 15 (1972): 1-12.
25. E. A. Lunzer, "Formal Reasoning: A Re-Appraisal," paper presented to the third annual meeting of the Jean Piaget Society, Temple University, May 1973.
26. R. Karplus, "Opportunities for Concrete and Formal Thinking on Science Tasks," delivered at the third annual meeting of the Jean Piaget Society, May 22, 1973, and to be published in the proceedings
27. Michael Shayer, "Some Aspects of the Strengths and Limitations of the Application of Piaget's Developmental Psychology to the Planning of Secondary School Science Courses," Master's Thesis, University of Leicester, 1972.
28. R. B. Ingle and M. Shayer, "Conceptual Demands in Nuffield O-Level Chemistry," Educ. in Chem., 8, 5, (Sept. 1971).
29. A. E. Lawson and W. Wollman, "Encouraging the Transition from Concrete to Formal Cognitive Functioning--An Experiment," Journal of Research in Science Teaching, submitted.
30. W. Wollman and A. E. Lawson, "Controlling Variables: Obstacles to be Overcome in the Acquisition of the Concept," School Science and Mathematics, submitted.
31. W. Wollman, "Intellectual Development Beyond Elementary School VII: Controlling Variables and Cognitive Conflict," School Science and Mathematics, submitted.

32. Antonio Suarez and Martin Rhonheimer, Lineare Funktion (Zürich: Limmat Stiftung, 1974).
33. R. Karplus and C. A. Lawson, SCIS Teacher's Handbook (Berkeley: University of California, 1974): 48-50.
34. F. P. Collea, R. G. Fuller, R. Karplus, L. G. Paldy, and J. W. Renner, "Workshop on Physics Teaching and the Development of Reasoning," Stony Brook: American Association of Physics Teachers, 1975.
35. A. E. Lawson and W. Wollman, "Physics Problems and the Process of Self-Regulation," The Physics Teacher, in press.

Mr. Short



Science class \_\_\_\_\_

Name \_\_\_\_\_

Grade \_\_\_\_\_

Age \_\_\_\_\_

Sex \_\_\_\_\_

What is the height of Mr. Tall, measured in paper clips? \_\_\_\_\_  
Please explain carefully how you found this answer.

---

---

---

---

---

---

---

---

Mr. Tall's car is 14 paper clips wide.  
How wide is Mr. Tall's car, measured in buttons? \_\_\_\_\_  
Please explain carefully how you found this answer.

---

---

---

---

---

---

---

---

APPENDIX A. Answer Page for Proportional Reasoning Task

APPENDIX A. Answer Page for Control of Variables Task

Name \_\_\_\_\_ Age \_\_\_\_\_ Grade \_\_\_\_\_

1. A sphere rolls down a ramp and hits a target. What might affect how far the target goes?

---



---



---



2. Suppose you want to know how much difference the weight of the sphere makes in how far the target goes. You are going to use the two spheres on the heavy target.

- (a) Where would you start the heavy sphere? \_\_\_\_\_  
 Where would you start the light sphere? \_\_\_\_\_  
 Please explain your answers carefully.

---



---



---

- (b) Suppose someone starts the light sphere from the low position. Where would you start the heavy sphere to find out how much difference its weight makes? \_\_\_\_\_

Please explain your answer carefully. \_\_\_\_\_

---

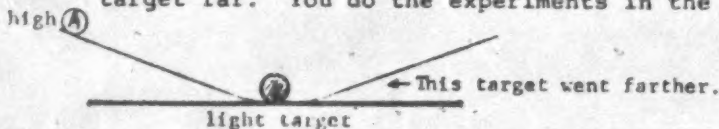


---



---

3. Suppose you have spheres of two different kinds of metal, metal A and metal B. You want to find out which metal is better for hitting a target far. You do the experiments in the two drawings.



What you see is that the metal A sphere hits its target much farther than the metal B sphere hits its target. Does this prove that metal A is better for hitting the target far? Yes \_\_\_\_\_ No \_\_\_\_\_

Please explain your answer in detail. \_\_\_\_\_

---



---



---

## APPENDIX B

In one of our earlier studies of proportional reasoning we found a substantial effect on the students' responses of the order of two questions, one of which was harder than the other.<sup>8</sup> In the present investigation, one may wonder to what extent a subject's first encounter with the "Mr. Tall" question may influence his approach to this and to the "car" question. We have already pointed out that students whose answers to the two questions differed usually gave a better answer to the "Mr. Tall" question. Yet Wollman pointed out to us that the larger numerical differences on the "car" question, if solved first, might well lead students to use a proportional rather than an additive approach.

To determine what influence the question order might have, we prepared an answer page on which the two questions were exchanged, and adapted the presentation of the task to the new order. This modified task was then administered to seventy-seven boys and ninety girls in the pilot group of eighth graders in the United States listed in Table 1. They were in two classes in an upper middle class community and in five classes in middle class communities in the same schools as students in the principal United States group.

The composite scores resulting from this investigation are presented in Table 11 for the pilot group as well as for a group of middle class and upper middle class students matched as closely as possible to the pilot group in ability levels and community. There is a statistically significant difference for boys ( $p < 0.05$ ) and for girls ( $p < 0.001$ ) between the two forms of the task. The reversed order leads to a substantial reduction in the Additive responses for both sexes, and a substantial increase in Ratio answers for boys, in Transitional answers for girls.

When the individual explanations to each of the two questions are examined, one finds that the largest decrease is in Additive responses to both questions and that there is an increase in papers with a Transitional answer to the "car" question and Ratio answer to the "Mr. Tall" question (composite Category Tr). For boys there is a further substantial increase in papers with ratio answers to both questions and a decrease in papers with Transitional answers to both questions.

Table 11 Composite Scores on the Proportional Reasoning Task for Reversed and Standard Presentation (percent)

|                       | (N)  | I  | A  | Tr | R  |
|-----------------------|------|----|----|----|----|
| Boys, Reversed Order  | (77) | 34 | 9  | 30 | 27 |
| Boys, Standard Order  | (77) | 29 | 27 | 27 | 17 |
| Girls, Reversed Order | (90) | 21 | 22 | 38 | 19 |
| Girls, Standard Order | (90) | 24 | 47 | 14 | 15 |

The attitudes of most classes differed when the reversed order was presented. After the question about the width of the car was posed, many students gasped at the task and were initially at a loss as to how to proceed. Some of them then answered the "Mr. Tall" question first, even though it was placed second on the answer page. The students perceived it as a very difficult assignment, yet many more ultimately applied transitional or proportional reasoning.

It is clear from this investigation that the "Mr. Tall" question elicits a particularly high number of Additive responses from students in the United States. The standard order of questions on the Proportional Reasoning task, with the "Mr. Tall" item first, therefore appears to be more challenging than the same task with the reversed order, and we used it for that reason. The fact that many student populations in our principal study nevertheless applied relatively little Additive reasoning is evidence of their good understanding of the concept of proportion. Some students in the early stages toward achieving proportional reasoning, however, were undoubtedly misled and thus their number was underestimated by our approach.